Czech University of Life Sciences

Faculty of Agrobiology, Food and Natural Resources

Department of Plant Protection

Identification and Characterization of Some Viruses Infecting Grapevine, Cucurbitaceous Vegetable and Ornamental Plants

Doctoral dissertation thesis

Author: Ing. Karima Ben Mansour, MSc.

Supervisor: Prof. Ing. Pavel Ryšánek, CSc.

Co-Supervisor: Ing. Petr Komínek, Ph.D.

Prague-2024

Declaration

I hereby declare that my doctoral dissertation thesis, entitled "Identification and Characterization of Some Viruses Infecting Grapevine, Cucurbitaceous Vegetable and Ornamental Plants," is the result of original investigation conducted under the supervision of Prof. Ing. Pavel Ryšánek, CSc., and Ing. Petr Komínek, Ph.D. Furthermore, I confirm that this dissertation has not been submitted elsewhere for any other degree or professional qualification, and I have properly acknowledged all the sources of materials used in this dissertation.

Prague 22/05/2024

When you want to succeed as bad as you want to breathe then you'll be successful

I believe in the person I want to become….

Dedication

Je dédie ce travail,

À mes très chers parents

Aucune dédicace ne saurait exprimer l'amour, l'estime, le dévouement et le respect que j'ai toujours eu pour vous.

Pour votre amour, votre patience et votre générosité, pour tous les efforts et les sacrifices que vous m'avez accordés. J'espère valoir votre affection et de votre confiance. Je vous dédie ce travail en témoignage de ma grande reconnaissance et de mon éternel amour. Que Dieu vous donne longue vie et bonne santé.

À mes adorables sœurs et mes frères, surtout Mohamed Ali

Qui n'a jamais cessé d'être là pour moi un modèle de persévérance, de courage et de générosité à suivre.

À mon fiancé Štěpán, et à tous les membres de ma famille, je leur souhaite tout le bonheur du monde.

À ces bonnes personnes, à qui je dois la réalisation de ce projet grâce à leur soutien et leur compréhension,

Au prof. Adrian J. Gibbs pour tous ses encouragements, son soutien et son aide auxquels prouve ma gratitude.

Je tiens particulièrement à exprimer ma reconnaissance envers Dr. Makovcová et Dr. Mazáková, Je vous remercie infiniment.

Acknowledgement

I would like to express my sincerest gratitude to my supervisors, Prof. Pavel Ryšánek, Dr. Petr Komínek, and Prof. Adrian J. Gibbs, for their invaluable guidance, academic support, and encouragement throughout my research. Their guidance and support were instrumental in the success of my work.

I would like to express my gratitude to Prof. Miloslav Zouhar, head of the Department of Plant Protection, for providing the financial support and facilities necessary to complete this work.

I would also like to extend my appreciation to the laboratory supervisors; Dr. Jana Mazáková and Mrs. Marcela Komínková for their guidance and help during my research.

Furthermore, I would like to express my gratitude to Dr. Thierry Wetzel and Dr. Patrick Winterhagen for their invaluable assistance during my one-year tenure in their laboratories. Furthermore, I would like to express my gratitude to Prof. Adrian J. Gibbs, Dr. Wulf Menzel and Dr. Paolo Margaria for their invaluable assistance and guidance in the field of bioinformatics.

I would also like to extend my gratitude to the Faculty of Agrobiology, Food and Natural Resources (FAFNR) staff and all those involved for making the faculty such a wonderful place to live and work. It was a great pleasure to interact with Dr. Kateřina Makovcová, and I am very thankful for her generous assistance and collaboration. I would like to extend my gratitude to the head of the dormitories, Mrs. Iveta Kuníková, for her assistance during my time in the dormitories.

Table of Contents

List of abbreviations

aa	Amino acid
BCE	Before Common Era
BCMV	Bean common mosaic virus
BEAST	Bayesian Evolutionary Analysis Sampling Trees
CE	Common Era
CMV	Cucumber mosaic virus
CP	Coat protein
CVA	Cherry virus A
ICTV	Internation Committee on Taxonomy of Viruses
GAMaV	Grapevine asteroid mosaic-associated virus
GFkV	Grapevine fleck virus
GLRaV-3	Grapevine leafroll-associated virus-3
GLRaV-4	Grapevine leafroll-associated virus-4
GPGV	Grapevine Pinot gris virus
GPLIV	Grapevine phloem-limited isometric virus
GRGV	Grapevine red globe virus
GRSPaV-1	Grapevine rupestris stem pitting-associated virus-1
GRVFV	Grapevine rupestris vein feathering virus
GSyV-1	Grapevine Syrah virus-1
GVA	Grapevine virus A
GVB	Grapevine virus B
kb	Kilobases
ML	Maximum Likelihood
MP	Movement protein
NGS	Next Generation Sequencing
NJ	Neighbour Joining
ORF	Open reading frame
PhyML	Phylogenetic Maximum Likelihood
PNRSV	Prunus necrotic ringspot virus

PPV	Plum pox virus
PRSV	Papaya ringspot virus
RAxML	Randomized Axelerated Maximum Likelihood
RdRp	RNA dependent RNA polymerase
RNA	Ribonucleic Acid
SbMV	Soybean mosaic virus
TMV	Tobacco mosaic virus
USA	United States of America
WMV	Watermelon mosaic virus
WSMV	Wheat streak mosaic virus

LIST OF FIGURES

Introduction

The genome of each plant virus only encodes a few proteins. They are obligate parasites that require a living host for replication and dissemination (Tatineni & Hein, 2023). Most plants harbour several viruses, and most viruses have a limited, and characteristic, host range. The symptoms caused by viruses in plants can vary considerably, depending on the virus and its host. Such disease symptoms may include mosaic, mottling, chlorosis, ringspots, distortion, dwarfing, and necrosis (Stevens, 1983). Consequently, viruses represent a significant risk to agricultural output. The impact of viruses on global plant production is significant as they damage food crops by decreasing their quality and quantity (Jones & Naidu, 2019). A total of more than 1,000 viruses have been identified as infecting plants (Gergerich & Dolja, 2006). A wide variety of agricultural crops, including cucurbitaceous vegetables, have been identified as susceptible to viral infection. In fact, more than 96 viruses have been identified as infecting these crops (Abdalla & Ali, 2021). Similarly, grapevines have been shown to be susceptible to around 100 viruses (Fuchs, 2020), and ornamental plants by more than 50 viruses (Mitrofanova *et al.*, 2018)

It is therefore of the utmost importance to identify and characterise viruses infecting plants in order to develop effective management strategies and ensure food security. To achieve this, a comprehensive analysis of the viruses in question is required, including an investigation into their origins, transmission patterns, and the impact they have on their host plants. It is also important, when possible, to reveal the past evolutionary history of each virus as this sometimes reveals the conditions under which it became an important pathogen, and this may indicate how best to control it.

Chapter I: Literature Review

I. Overview of the affected plants

1. Pumpkin

Vegetables are of global economic and horticultural importance. One of the largest vegetable families is the *Cucurbitaceae*, which contains approximately 115 genera and 960 species (Schaefer & Renner, 2011). Some of the most important edible plants in the world are members of this family (Guo *et al.*, 2020). The members of the *Cucurbitaceae* family, commonly referred to as 'cucurbits' or 'gourds', can be categorised into three main groups: annual herbaceous vines, perennial tubers or roots, and perennial lianas (Guo *et al.*, 2020). These plants originate from tropical, subtropical, and temperate regions. The cucurbits include melon (*Cucumis melo)*, cucumber *(Cucumis sativus)*, watermelon (*Citrullus lanatus*), zucchini (*Cucurbita pepo* var. *cylindra*), and pumpkin (*Cucurbita pepo* var. *pepo*).

It is important to note that the term 'pumpkin' refers to several cultivated species, including *Cucurbita pepo* L., which is the most grown and economically important (Ferriol & Picó, 2008), and also *Cucurbita maxima* Duchesne, *Cucurbita moschata* Duchesne, *Cucurbita argyrosperma* Huber, and *Cucurbita ficifolia* Bouché (Provesi & Amante, 2015).

The domestication of cucurbits was not a uniform process, with the domestication of melon (*Cucumis melo*) and cucumber (*Cucumis sativus*) occurring in Asia and dated to more than 4000 years ago (Paris, 2016). In contrast, the domestication of watermelon (*Citrullus lanatus*) occurred in Africa (Egypt, Libya, and Sudan) between 4000 and 5000 years ago (Renner *et al.*, 2021). Finally, zucchini (*Cucurbita pepo* var. *cylindra*) and pumpkin (*Cucurbita pepo* var. *pepo*) were domesticated in the Americas 10,000 years ago (Chomicki *et al.*, 2020).

Abiotic and biotic factors exert a profound influence on the production and quality of cucurbit fruits. Among the biotic factors, viruses represent a significant concern as approximately 50% of emerging infectious diseases of cucurbits were attributed to viruses (Anderson *et al.*, 2004) including 59 viruses (Lecoq & Desbiez, 2012), but that number almost doubled in a decade (Abdalla & Ali, 2021). The major viruses infecting cucurbits do not belong to the same genus, and they can exist in multiple and mixed infections. Some of the viruses have synergistic relationships that can enhance the severity of symptoms and, therefore, decrease the yield more (Lecoq & Desbiez, 2012).

2. Grapevine

The genus *Vitis* comprises approximately 60 to 80 species of vines (Hardie, 2000; Emanuelli *et al.*, 2013). The most widely cultivated grapevine species is *Vitis vinifera* subsp. *vinifera* which was domesticated approximately 11,000 to 7,000 years ago (Myles *et al.*, 2011; Allaby, 2023; Dong *et al.*, 2023). It is a perennial woody crop (Zhu *et al.*, 2021), that produces clusters of small, round fruits called grapes, which can be red, white, yellow, pink, crimson, dark blue, or black in colour (Naidu *et al.*, 2014). Grapes are mainly used for winemaking. The different grape varieties and winemaking techniques result in a diverse range of wine styles and flavours (Chambers & Pretorius, 2010; González-Neves *et al.*, 2013; Morata *et al.*, 2017). Additionally, the fruit can be consumed or dried for the production of juice, raisins, and vinegar (Verter & Hasíková, 2019). It is widely cultivated for its significant economic value worldwide (Zhu *et al.*, 2021). Grapevines are susceptible to a variety of pests and infections especially as they are propagated vegetatively, which allows pathogens to accumulate (Naidu *et al.*, 2014). The most damaging of these are viruses and virus-like entities (Naidu *et al.*, 2014). A total of more than 100 viruses have been identified as infecting grapevines (Fuchs, 2020; Tatineni & Hein, 2023). These viruses can result in yield losses, a reduction in the productivity of the vineyard, and reduce the lifespan of the vine.

3. Myrobalan plum

The myrobalan plum, also known as the cherry plum (*Prunus cerasifera*), is a small shrubby tree (Popescu & Caudullo, 2016) native to Southeast Europe, Western and Middle Asia (Popescu & Caudullo, 2016). The flowers of *Prunus cerasifera* are of importance to pollinators (Petrov *et al.*, 2024). It is an easily adaptable species (Czortek *et al.*, 2024). *Prunus cerasifera* produces plum-like edible fruit (Popescu & Caudullo, 2016; Petrov *et al.*, 2024). *Prunus cerasifera* is found to be one of the parents of the cultivated plum, *Prunus domestica* (Zohary, 1992). It is commonly used as an ornamental plant (Petrov *et al.*, 2024), but it is highly valued as a grafting stock for other *Prunus* species and cultivars. This is due to its resistance to frost, root-knot nematodes, and its ability to improve fruit weight, greater tolerance to the infection by being symptomless or showing mild spots, or as a biological indicator for the viruses (Southwick *et al.*, 1999; Milliron *et al.*, 2021). Together with other *Prunus* species, the myrobalan plum serves as a natural host for numerous stone fruit viruses, acting as a reservoir for these viruses (Kamenova, 2008; Gospodaryk *et al.*, 2013).

II. Viruses studied in this Ph.D. work

1. Watermelon mosaic virus

The virus belongs to the *Potyvirus* genus and has a single positive-stranded RNA. It is an important pathogen of cucurbits (Lecoq & Desbiez, 2012). The symptoms of watermelon mosaic virus (WMV) can be observed on leaves through mosaic, vein banding, or on fruits through discolouration, deformation, and size reduction (Figure 1).

Figure 1. Symptoms of the watermelon mosaic virus infection (©Crop Research Institute, Prague - Ruzyne)

This virus was reported to be the progeny of an interspecific recombination between two legume-infecting viruses: soybean mosaic virus (SbMV) and bean common mosaic virus (BCMV) (Desbiez & Lecoq, 2004). The WMV has a broad host range. It can infect other plants including legumes, orchids, and weeds (Desbiez *et al.*, 2007; Lecoq & Desbiez, 2012; Gao *et al.*, 2021). In experimental conditions, WMV has been demonstrated to infect 170 plant species (Wang & Li, 2017). It is present worldwide, including the Czech Republic (Svoboda, 2011). WMV is transmitted by aphids in a non-persistent manner (Lecoq & Desbiez, 2012). A total of 35 aphid species have been identified as vectors for WMV (Wang & Li, 2017). Of these, three aphid species, namely (*Aphis craccivora, Aphis gossypii*, and *Myzus persicae*) have been identified as particularly efficient vectors of WMV (Lecoq & Desbiez, 2012).

2. Grapevine Pinot gris

The grapevine Pinot gris virus (GPGV) was first reported in Italy in 2012 and is a relative newcomer to the various pathogens affecting viticulture worldwide (Giampetruzzi *et al.*, 2012). It is a member of the *Trichovirus* genus of the *Betaflexiviridae* family, which consists of a positive single-stranded RNA with three overlapping open reading frames encoding replicase (RdRp), movement protein (MP), and coat protein (CP). The virus was isolated from asymptomatic and symptomatic grapevines, and the symptoms ranged from mild up to severe, causing grapevine stunting and yield loss. The presence of symptoms has been reported to vary according to GPGV isolates and also according to cultivar or seasonal environmental parameters (Tarquini *et al.*, 2019b, 2021). One probable vector of the virus is the eriophyid mite *Colomerus vitis*. Nevertheless, the observed dissemination pattern in vineyards and the documented occurrence of GPGV in non-*Vitis* host plants suggest the potential involvement of additional animal vectors (Malagnini *et al.*, 2016; Gualandri *et al.*, 2017; Hily *et al.*, 2021; Demian *et al.*, 2022). Furthermore, the ease with which GPGV can be distributed between countries and continents through the exchange of infected plant materials underscores the necessity for heightened vigilance. It is possible that the virus entered Eastern Europe prior to 2005 and subsequently spread to other European countries after 2010 (Bertazzon *et al.*, 2016). In parallel, GPGV has been reported from many grapevine-growing regions across different countries worldwide (Giampetruzzi *et al.*, 2012; Glasa *et al.*, 2014; Beuve *et al.*, 2015a; Jo *et al.*, 2015; Reynard *et al.*, 2016; Rwahnih *et al.*, 2016; Ruiz-García & Olmos, 2017; Czotter *et al.*, 2018; Rasool *et al.*, 2019; Zamorano *et al.*, 2019; Debat *et al.*, 2020; Eichmeier *et al.*, 2020; Massart *et al.*, 2020; Abe & Nabeshima, 2021; Navrotskaya *et al.*, 2021).

3. Grapevine fleck virus

The virus was first isolated in 1983 in the phloem of infected vines and named the grapevine phloem-limited isometric virus (GPLIV) (Castellano *et al.*, 1983). In 1991, it was renamed as the grapevine fleck virus (GFkV) in recognition of its role as agent responsible for fleck (Boscia *et al.*, 1991). Its genome is single-stranded positive-sense RNA of 7.5 kb in length (Sabanadzovic *et al.*, 2000). The Latinized binomial name for the taxon to which it belongs is *Maculavirus vitis* species (International Committee on Taxonomy of Viruses; Walker *et al.*, 2021; Zerbini *et al.*, 2022). The virus is phloem-limited (Shi *et al.*, 2003). The symptoms generated by this virus differs in different *Vitis* species. The virus causes the clearing of the veinlets in young leaves of *Vitis rupestris* Scheele, and distortion in older leaves. However, it is latent in *Vitis vinifera* L. (Martelli

et al., 2002). To date, no vector for the virus has been identified, but it can be transmitted via grafting and dispersed by the exchange of grapevine propagation materials (Martelli, 2014). The virus was reported from many grapevine growing regions (Komínek & Holleinová, 2003; Jo *et al.*, 2017; Sabanadzovic *et al.*, 2017; Crnogorac *et al.*, 2020).

4. Grapevine red globe virus

The grapevine red globe virus (GRGV) is a member of the *Maculavirus* genus. It also induces specific symptoms in *Vitis rupestris* Scheele but does not cause any symptoms in *Vitis vinifera* L.(Sabanadzovic *et al.*, 2000; Ghanem-Sabanadzovic *et al.*, 2003). The virus was first reported in grapevines from Italy and Albania (Sabanadzovic *et al.*, 2000) and has since been documented in many other European countries, including Greece (El Beaino *et al.*, 2001), France (Beuve *et al.*, 2015b), Spain (Cretazzo *et al.*, 2017), Hungary (Czotter *et al.*, 2018), Germany (Ruiz-García *et al.*, 2018), Czech Republic (Massart *et al.*, 2019), Portugal (Candresse *et al.*, 2022), and United Kingdom (Dixon *et al.*, 2022) and non-European countries (USA (El Beaino *et al.*, 2001), China (Fan *et al.*, 2016), Brazil (Fajardo *et al.*, 2017), Iran (Nourinejhad Zarghani *et al.*, 2021), Japan (Yamamoto *et al.*, 2022), and Australia (Wu *et al.*, 2023) were also affected.

5. Grapevine rupestris vein feathering virus

Grapevine rupestris vein feathering virus (GRVFV), is a member of the genus *Marafivirus*. The virus has a positive-sense, single-stranded RNA genome. The genome encodes a large ORF that produces a putative polyprotein comprising methyltransferase, peptidase, helicase, RNA polymerase, and coat protein domains. The virus was first reported in Greece (El Beaino *et al.*, 2001) and has since been identified in many grapevine-growing regions worldwide (Eichmeier *et al.*, 2016; Wu *et al.*, 2021; Shvets *et al.*, 2022). The virus causes mild asteroid symptoms in *Vitis vinifera* L. and vein feathering in *Vitis rupestris* Scheele (El Beaino *et al.*, 2001). Infection can occur alone or in combination with other grapevine viruses such as GRGV (Fiore *et al.*, 2016) and with Grapevine Syrah virus-1 (Miljanić *et al.*, 2022).

6. Grapevine Syrah virus-1

Grapevine Syrah virus-1 (GSyV-1) is a member of the *Marafivirus* genus, within the species *Marafivirus syrahensis*. The genome is 6.5 kb in length, and it encodes a large ORF that produces a 2081 aa polyprotein. The virus was first reported from the USA in *Vitis vinifera* cv Syrah (Al Rwahnih *et al.*, 2009). Since then, the virus has been found in Chile (Engel *et al.*, 2010),

Italy (Giampetruzzi *et al.*, 2012), Hungary (Czotter *et al.*, 2015), South Africa (Oosthuizen *et al.*, 2016), China (Ahmed *et al.*, 2018), Croatia (Vončina *et al.*, 2017), Spain (Ruiz-García *et al.*, 2017), Korea (Cho *et al.*, 2019), Russia (Navrotskaya *et al.*, 2021), Czech Republic and Slovakia (Glasa *et al.*, 2015), and many other countries. There is limited data on the impact of GSyV-1 on grapevine production (Vončina *et al.*, 2017; Shvets *et al.*, 2022). GSyV-1 is not limited to infecting *Vitis vinifera*, as it has also been found in wild blackberries, where it was named Grapevine virus Q (Sabanadzovic *et al.*, 2009).

7. Grapevine asteroid mosaic-associated virus

Grapevine asteroid mosaic-associated virus (GAMaV) is a member of the genus *Marafivirus*. It is thought to be transmitted through bud-wood (Sabanadzovic *et al.*, 2017). Furthermore, no vector associated with virus transmission has been identified, although, seed transmission is a potential mode of transmission (Thompson *et al.*, 2021). The virus has been reported in many countries, including USA (Al Rwahnih *et al.*, 2009), Canada (Xiao & Meng, 2016), Japan (Nakaune *et al.*, 2008), Uruguay (Jo *et al.*, 2015), France (Candresse *et al.*, 2017), Spain (Morán *et al.*, 2020), Italy (Porceddu *et al.*, 2018), Russia (Shvets *et al.*, 2022), and Hungary (Czotter *et al.*, 2018). The symptoms manifest as an asteroid mosaic described as "star-shaped chlorotic spots" on the grapevine leaves (Thompson *et al.*, 2021). Recently, grapevines infected with this virus and showing chlorotic mottling have been found in China and Spain (Morán *et al.*, 2020; Fan *et al.*, 2023). Furthermore, the virus has also been identified in asymptomatic samples (Morán *et al.*, 2020; Fan *et al.*, 2023).

8. Prunus necrotic ringspot virus

Prunus necrotic ringspot virus (PNRSV) is a member of the genus *Ilarvirus* and is known to infect both *Prunus* spp. and ornamental plants (Pallas *et al.*, 2013). The genome is segmented, and tripartite: RNA1 and RNA2 encode two replicase proteins, P1 and P2, respectively, while RNA 3 encodes two additional proteins: the movement protein (MP) and coat protein (CP) (Pallas *et al.*, 2013). The virus is distributed globally. In the Czech Republic, the virus was first identified serologically in sour and sweet cherries (Karešová *et al.*, 1986). PNRSV can be transmitted by pollen, which facilitates rapid virus spread in orchards (Kryczynski *et al.*, 1992), or by seed. The efficiency of these two natural modes of transmission varies depending on the host plant species (Barba *et al.*, 1986). Additionally, PNRSV can be transmitted by infected plant-propagating

material, such as budwood and rootstocks. The initial symptoms of PNRSV manifest one-year post-infection, designated the acute or shock stage. However, subsequent to this, the plants become symptomless. Nevertheless, previous studies have indicated that certain strains may result in recurrent symptoms each year (Nyland *et al.*, 1976; Wells & Kirkpatrick, 1986). The symptoms exhibited by infected plants are dependent on the specific isolate of PNRSV, including mosaic, ringspot, chlorosis, leaf deformation, necrosis, shot holes, and drop-off. Moreover, it has been demonstrated that the virus significantly affects fruit yield and quality.

9. Cherry virus A

Cherry virus A (CVA) is classified within the genus *Capillovirus* (Jelkmann, 1995). The genome consists of a positive single-stranded RNA with two ORFs; ORF1 encodes a replicase and coat protein, while ORF2 encodes a movement protein in a different frame (Jelkmann, 1995). The symptoms caused by this virus have been reported as latent or unknown due to the fact that the virus is usually found in mixed infections, which makes it difficult to associate the virus with specific disease symptoms (Gao *et al.*, 2016). This virus is distributed worldwide. The virus was first reported in 2010 in sweet and sour cherries in the Czech Republic (Grimová et al, 2010). CVA is transmitted via grafting; however, vector transmission has not yet been reported. The virus has also been identified in non-cherry prune hosts, including apricot, plum, peach, and Japanese apricot (Kesanakurti *et al.*, 2017).

III. Evolution of viruses

A. Evolution and phylogenetics

Phylogenetics is the study of the evolutionary history of species, which concerns the relationship between lineages that are related by a common ancestor (Gorbalenya & Lauber, 2017).

This branch of science flourished almost two centuries ago a result of the insights of Charles Darwin (Anon, 2024a) and Alfred Wallace (Anon, 2024b). Darwin summarized the possibility that each living organism evolves and produces a bifurcating lineage (i.e. phylogeny) in the form of a tree, shown in his famous sketch (Figure 2).

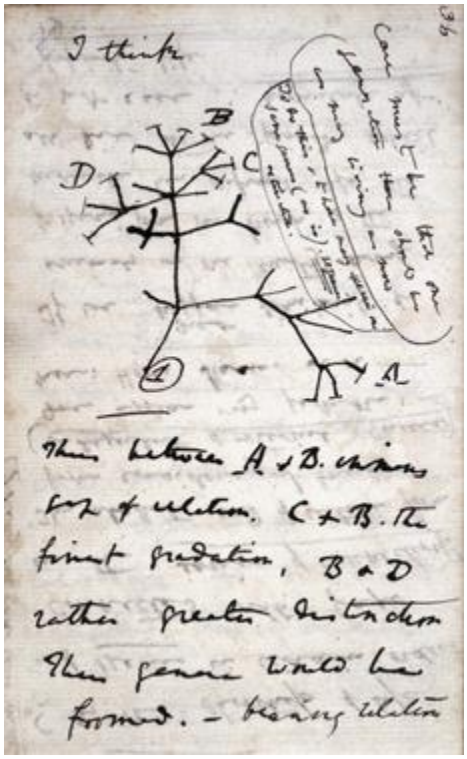

Figure 2. Page from Darwin's notebooks (c. July 1837) with his first sketch of an evolutionary tree, and the words "I think" at the top. Retrieved from (Anon, 2024a)

Initially phylogenies were inferred by observing and comparing superficial characters of organisms. However the sequences of, first, amino acids in proteins, and then nucleotides in nucleic acids of organisms (Dayhoff, 1978) were shown to change through time in a similar way

(Simmons *et al.*, 2002), and it became possible to calculate with more accuracy phylogenies representing the evolutionary history of each organism (Gorbalenya & Lauber, 2017).

The construction of a phylogenetic tree is a complex process that involves many steps (Choudhuri, 2014). The initial step is to select the molecular sequence dataset, which can be either nucleotides in nucleic acids or amino acids in proteins. The phylogenetic tree can be constructed using either genomes or distinct genes. The next step is to make a multiple sequence alignment, which serves to maximize similarities and to identify mutations and other evolutionary changes, such as recombination or reassortment (shuffling) of parts of the sequence. This is followed by the selection of an appropriate model of evolution, and the construction of the phylogenetic tree. Finally, the topology of the tree is assessed using bootstrapping (Choudhuri, 2014) or other statistical estimate of likelihood (Shimodaira & Hasegawa, 1999). The outcome is a graphical representation of the evolutionary relationships of the population of sequences (Bawono & Heringa, 2014). The phylogenetic tree is described by its topology, namely its branches, their length, shape, positions and density of the nodes, and the position of the root (Gorbalenya & Lauber, 2017).

Phylogenetic tree comparisons may be employed to classify novel viruses or an isolates exhibiting novel or unexpected properties (Burrell *et al.*, 2017). Moreover, the construction of phylogenetic trees has been employed not only as a tool for understanding isolates with new or unexpected properties, but also for a more accurate taxonomic classification of these isolates. The International Committee on Taxonomy of Viruses (ICTV) has recently accepted a proposal to utilise phylogenetic analyses as the primary criterion for classifying the multitude of newly identified viruses and metagenomic sequences that have been obtained through next generation sequencing (NGS) technique, without the necessity to include biological information, given the unfeasibility of gathering such information due to the growing number of newly identified viruses and metagenomes (Simmonds *et al.*, 2017); a change opposed by some (Gibbs, 2020).

Evolution refers to the process by which viruses undergo change over many generations and periods of time. Once these changes, or mutations, are fixed, a distinctive lineage is formed (Gorbalenya & Lauber, 2017). Therefore, the objective of phylogenetics is to shed light on the manner in which viruses have evolved and to make predictions about their shared history (Pagán, 2018). There are many different computational tools that can be used to transform the information about the detected mutations into a phylogenetic tree. These include distance-based methods such

as neighbour-joining, which can be computed relatively quickly, and evolutionary-based methods, such as maximum likelihood or Bayesian (Pappas *et al.*, 2020), some of which take longer. Further statistical and mathematical models of evolution can be found in software packages such as Randomized Axelerated Maximum Likelihood (RAxML), Bayesian Evolutionary Analysis Sampling Trees (BEAST), and Phylogenetic Maximum Likelihood (PhyML), which provide estimates of substitution rates, divergence times, and other population genetics patterns (Pappas *et al.*, 2020).

Interpretation of phylogenies is aided by comparing the results of using completely different methods of calculation, for example, 'neighbour-joining' and 'maximum likelihood', as different models and algorithms are used. Interpretation of a phylogeny is also greatly aided by dating it, and again a great variety of methods are available. Dating of measurably evolving viruses is usually done using 'tip dating', namely using sequences from samples collected on different occasions, and then, if the dates and positions of the tips are correlated (i.e. there is a 'temporal signal'), then it is possible to extrapolate and date of the tree. Alternatively, for sequences that are so distantly related that there is no 'temporal signal' and extrapolations are too large then external calibration dates are required, which may include recognised evolutionary events or, ultimately, fossil records and radioactive decay (Pappas *et al.*, 2020). A dated phylogeny may then produce useful insights about factors affecting the spread of a virus. For example, Hajizadeh *et al.* (2019) reported that the phylogeny of sequences of plum pox virus (PPV) had a root that was probably about the beginning of the first millennium BCE, which, interestingly, did not coincide with the domestication of woody *Prunus* species, several thousand years ago, but did coincide with the invention of pruning for propagating species that do not root easily from cuttings, such as apples, pears, and plums.

Another study investigated the global origin and source of the Australian epidemic of wheat streak mosaic virus (WSMV) found that the virus, and its relatives, likely originated in Eurasia, especially Iran, and subsequently spread to other regions, including North America, and thence to Australia, and South America in plant breeders stocks, resulting in the Australian WSMV epidemic (Jones *et al.*, 2022). This emphasised the importance of national quarantine practices.

B. Evolution and genetic diversity

Variability is a fundamental characteristic of living entities. The genetic structure of a viral population is the distribution of genetic variants in the viral population. The process through which viruses change across many generations is referred to as evolution (García-Arenal *et al.*, 2001). The evolution of virus is driven by many processes such as mutation, genetic exchange through recombination or reassortment, which induce viral diversification and generation of novel viral strains, which may affect host range, viral pathogenicity, and transmission (Escriu, 2017).

1. Sources of genetic diversity

1.1.Mutation

RNA viruses possess an RNA-dependent RNA polymerase (RdRp), which lacks proofreading activity, resulting in the generation of many genetically distinct sequences within viral populations (Drake & Holland, 1999), called a "quasi-species". The mutation rate of RNA virus genomes has been repeatedly estimated to be between 10^{-3} and 10^{-4} for some viruses, including tobacco mosaic virus (TMV) (Malpica *et al.*, 2002), cucumber mosaic virus (CMV) (Ouedraogo & Roossinck, 2018), and papaya ringspot virus (PRSV) (Khanal & Ali, 2021).

1.2.Recombination

It is defined as the process of shuffling viral RNA segments of two or more related viruses (MacFarlane, 1997). Recombination events only occur within a single multiply-infected host plant, and contribute to the genetic diversity of viral populations (Nagy, 2008). Plant viral recombination and reassortment play an important role in the evolution and adaptation of plant viruses (Moreno *et al.*, 2004). It increases the adaptability of viruses to new host ranges (Maliogka *et al.*, 2012), the emergence of new viral strains, an increase in virulence and pathogenicity (LaTourrette & Garcia-Ruiz, 2022), changes in the specificity of transmission vectors (Gadhave *et al.*, 2020), and resistance to control measures (Moreno *et al.*, 2004; Pérez-losada *et al.*, 2015; Gibbs *et al.*, 2020). The process of RNA recombination is not necessarily reciprocal (Pérez-losada *et al.*, 2015), and is thought to occur during viral replication, when the viral replicase enzyme switches between different RNA templates, leading to the formation of chimeric sequences from parental genomes (Nagy, 2008).

A substantial number of plant viruses undergo RNA recombination, as evidenced by studies on potyviruses. Plum pox virus (Glasa *et al.*, 2004), potato virus Y (Green *et al.*, 2017), tulip breaking

virus (Ágoston *et al.*, 2020), and watermelon mosaic virus (Desbiez & Lecoq, 2008; Desbiez *et al.*, 2011). These are just a few of many reported examples of RNA recombination in potyviruses.

It has been documented that other viruses infecting perennial hosts have been identified to undergo recombination, including viruses infecting grapevine. This process has been observed in grapevine Pinot gris virus (Tarquini *et al.*, 2019a; Hily *et al.*, 2020), grapevine leafroll-associated virus-3 (GLRaV-3), grapevine leafroll-associated virus-4 (GLRaV-4), grapevine rupestris stem pitting-associated virus-1 (GRSPaV-1), grapevine virus A (GVA), grapevine virus B (GVB), and GSyV-1 (Fajardo *et al.*, 2017).

It has been reported that viruses infecting stone trees such as cherry virus A and prunus necrotic ringspot virus are recombinants (Boulila, 2010; Gao *et al.*, 2017).

2. Evolutionary forces that shape the genetic diversity

A fundamental concept in the study of evolution is the survival of the fittest genetic variants within a given population. The concept of fitness is defined as the capacity of a genetic variant, to contribute to the next generation relative to that of other molecular variants in the population, in specific environmental conditions (Maynard Smith, 1998). Therefore, the estimate of fitness of this genetic variant is the frequency at equilibrium with which this genetic variant is present in the progeny. Two forces, namely selection and genetic drift are responsible for shaping of this fitness (Escriu, 2017).

2.1.Selection

The process of selection is defined as the alteration in the frequency of specific molecular variants within a population in a specific environment. In the context of evolutionary biology, positive or adaptive selection occurs when the most fit molecular variants increase in frequency within a population. Conversely, negative or purifying selection occurs when the least fit molecular variants decrease in frequency within a population (Pagán, 2018).

2.2.Genetic drift

As previously stated, viruses exhibit high rates of replication and mutation, which collectively result in the formation of large population sizes (Hughes, 2009). Nevertheless, during the process of infection or plant-to-plant transmission, viral populations may undergo a reduction in effective population size (or size of an idealized population) due to the occurrence of bottlenecks (Moury *et*

al., 2006). The threshold of effective population size was estimated to be about 10 (French & Stenger, 2003). In the event that the population size is insufficiently large, the transmission of genetic variants from one generation to the next may occur randomly, which is referred to as genetic drift (Escriu, 2017), or even to a collapse of one of the components of mixed population as a result of 'Muller's Ratchet' or a mutational meltdown (Fraile *et al.*, 1997).

Chapter 2: Hypotheses and objectives

1. Dissertation hypotheses

- The adaptation of plant viruses to new host species can be elucidated by examining historical genetic evidence.

- Viruses infecting trees exhibit distinct genetic characteristics and distribution patterns.

2. Dissertation objectives

-To determine the recent and historical origin of WMV and to analyse its spread.

- To update the historical origin of the introduction of GPGV.

- To investigate the prevalence and impact of multiple infections with viruses of the family *Tymoviridae*.

- To characterize the genetic variability of prunus necrotic ringspot virus and cherry virus A in myrobalan rootstock and their effects on the host plant.

Chapter 3: List of published studies

I. Study 1

-Title: Watermelon mosaic virus in the Czech Republic, its recent and historical origins

-Brief description: Watermelon mosaic virus (WMV) is a potyvirus and a member of the bean common mosaic virus (BCMV) lineage. The initial report of Czech WMV in 2011 was based solely on ELISA testing, which merely identifies the virus. The present study therefore aimed to place Czech isolates in the species molecular grouping based on gene sequences for the first time, to describe the diversity of Czech WMV, and to use all the available NCBI data of the full coding region of WMV sequences to provide information on the source of the Czech WMV population and the origin of WMV. Comparative dating indicates that the basal Chinese isolates are descendants of a potyvirus population infecting various dicotyledonous plant species in China at least 2000 years ago. WMV became a crop pathogen around 1000 years ago, a few years after watermelon was introduced to northern China and first cultivated as a crop during the Five Dynasties (907–960 CE).

Received: 4 April 2023 | Accepted: 29 May 2023

DOI: 10.1111/ppa.13766

Plant Pathology WILEY

ORIGINAL ARTICLE

Watermelon mosaic virus in the Czech Republic, its recent and historical origins

Karima Ben Mansour[1,2] | Adrian J. Gibbs[3] | Marcela Komínková[2] | Petr Komínek[2] | Jana Brožová[2] | Jan Kazda[1] | Miloslav Zouhar[1] | Pavel Ryšánek[1]

[1]Department of Plant Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Prague, Czech Republic

[2]Crop Research Institute, Prague, Czech Republic

[3]Emeritus Faculty, Australian National University, Canberra, Australia

Correspondence
Karima Ben Mansour, Department of Plant Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Czech Republic.
Email: karina79@hotmail.fr

Abstract

Watermelon mosaic virus (WMV) is a potyvirus and a member of the bean common mosaic virus (BCMV) lineage. It is one of the most economically important viral pathogens of cucurbits worldwide and was first reported in the Czech Republic in 2011 from serological surveys (2005–2011). In this study, we confirmed this identification by determining the complete coding regions of five Czech WMV isolates using high-throughput sequencing and Sanger sequencing (MW188031; OP585149–OP585152), together with the coat protein (CP) genes of 26 additional isolates. Phylogenies were made from these and more than 128 genomes or 128 CP genes from GenBank. They showed that the Czech isolates were most closely related to other European isolates, but, surprisingly, 96.2% of the genomes were recombinant. The nonrecombinant sequences mostly came from basal isolates, all originating from China, and some from unusual hosts (*Ailanthus altissima*, *Alcea rosea* and *Panax ginseng*). The complete WMV genomes form three phylogenetic clades, two of them small and basal, and the third includes all other isolates. Comparative dating suggests that the basal Chinese isolates are descendants of a potyvirus population infecting various dicotyledonous plant species in China at least 2000 years ago. WMV became a crop pathogen around 1000 years ago, a few years after watermelon was taken to northern China and first grown as a crop during the Five Dynasties (907–960 CE).

KEYWORDS
comparative dating, phylogenetic analysis, virus recombination, WMV

1 | INTRODUCTION

Potyviruses are among the largest genera of RNA viruses that infect plants (Inoue-Nagata et al., 2022). Potyviruses have flexuous filamentous virions measuring 650–950 nm in length and 11–12 nm in diameter. These virions contain a monopartite, single-stranded, positive-sense RNA genome of 9.7–11 kb, which is 3′ polyadenylated. The genome contains one large open reading frame (ORF), which is translated into a polyprotein that is autocatalytically hydrolysed into 11 proteins (Inoue-Nagata et al., 2022) including a protein denoted Pretty Interesting Potyviridae ORF (PIPO; Chung et al., 2008). Potyviruses infect a wide range of plant species, both monocotyledonous and dicotyledonous. They are spread by aphids in a nonpersistent manner, as well as sometimes by seeds and also by the human exchange and movement of plant materials (summarized by Gibbs et al., 2020).

One of the most important potyviruses is watermelon mosaic virus (WMV), first reported in Israel (Cohen & Nitzany, 1963) and then in many countries worldwide. WMV infects cucurbits and significantly affects the quality and quantity of their produce worldwide (Desbiez & Lecoq, 2008). It can also infect non-cucurbits, such

as peas, orchids, and, in total, at least 170 species from 27 families (Desbiez et al., 2007; Gao et al., 2021; Wang et al., 2017).

In the Czechia, cucurbit vegetables, which are the primary field hosts of WMV, are grown in the warmest regions of the country located in the lowlands of Bohemia in the west and Moravia in the east. The presence of WMV in squash and melon plants was first reported in the Czech Republic by Svoboda (2011). During surveys, conducted mainly from 2005 to 2010, WMV was serologically detected only in the vegetable-growing regions in southern and central Moravia, and not in Bohemia.

The intraspecies variability of WMV isolates has been reported and studies showed that the isolates then known could be placed in three molecular groups: G1 or classic (CL) isolates, group 2 (G2), and group 3 (G3), known as emerging isolates (EM), which were further divided into four subgroups, EM1–EM4 (Desbiez et al., 2007, 2009, 2011).

WMV belongs to the bean common mosaic virus (BCMV) lineage of potyviruses, which emerged several thousand years ago and diverged first in South-east and East Asia (Gibbs et al., 2008). Furthermore, early studies reported that WMV originated from interspecific recombination between two legume-infecting viruses: soybean mosaic virus (SbMV) and BCMV (Desbiez & Lecoq, 2004). Recombination and mutation are important sources of genetic variation and are the main factors in plant virus evolution (García-Arenal et al., 2001), leading to the generation of new strains (Desbiez & Lecoq, 2004), and changes of viral virulence (Chare & Holmes, 2006; Varsani et al., 2008). For potyviruses, RNA recombination appears to be important for their adaptation to new hosts and changing environmental conditions (Gibbs et al., 2020).

Phylogenies are interpreted more easily and informatively if they are dated, especially if features of the phylogenies are then found to coincide with historical events, thereby indicating the probable accuracy of the dating. The techniques most often used for dating phylogenies of organisms such as potyviruses, which evolve at measurable rates, use the temporal signal provided by gene sequences collected on different dates (i.e., heterochronous samples). However, the presence of many recombinants in the WMV population rules out the use of these techniques (Gao et al., 2020). Fortunately, the world populations of two closely related viruses, potato virus Y (PVY) and potato virus A (PVA), have been convincingly dated (Fuentes et al., 2019; Gibbs et al., 2017), and therefore, the "subtrees dating" method can be used to date phylogenies of WMV (Hajizadeh et al., 2019; Mohammadi et al., 2018). This is based on the fact that a single maximum-likelihood (ML) phylogeny describes the evolutionary distances between all the sequences in the phylogeny using a single statistical model; thus, distances in different parts of a single ML tree can, with caution, be directly compared with each other.

The first report of Czech WMV by Svoboda (2011) was based only on ELISA testing, which merely identifies the virus. Therefore, the present study aimed to place, for the first time, Czech isolates in the species molecular grouping based on gene sequences, to describe the Czech WMV diversity, and to use all the available NCBI data of the full coding region of WMV sequences to provide information on the source of the Czech WMV population and the origin of WMV.

2 | MATERIALS AND METHODS

2.1 | Sample collection, ELISA, reverse transcription PCR and sequencing

Pumpkin plants showing severe symptoms of WMV infection were observed in 2019 in a private garden in the small town of Smecno, Bohemia, and therefore, the isolate of WMV that was obtained was named WMV-Smecno. This isolate was deposited in the collection of plant viruses maintained at the Crop Research Institute, Prague (accession VURV-V 28.6, Culture Collection of Microorganisms of Crop Research Institute; acronym: VURV; World Data Center for Microorganisms reg. no.: 1236). In the collection, the virus is artificially propagated in zucchini plants; the collection also contains more WMV isolates, which were also used in this study.

As WMV had previously been found only in the Moravia region of Czechia, a larger survey was done in the vegetable-growing regions of Bohemia. Pumpkin leaves showing symptoms were collected and immediately tested using serological methods. A commercial polyclonal antibody (DSMZ, Braunschweig, Germany) was used for virus detection. The positive control was purchased from DSMZ and confirmed the accuracy of the results. Cucumbers and zucchini leaves grown in a glasshouse were used as the negative controls.

Among the ELISA-positive samples, 27 samples were randomly chosen for further molecular analysis. Two Moravian isolates, maintained in the VURV Collection, were also used. Total RNA was isolated from 100 mg of fresh leaf tissue using the Spectrum Plant Total RNA Kit (Sigma-Aldrich). One-step reverse transcription (RT)-PCR was performed, using primers listed in Table S1, to amplify different regions of the genome; these were assembled using Clone Manager 9 to obtain four almost full-length sequences. Primers WMV-8180-F and WMV-10040-R (Table S1) were used to amplify the coat protein (CP) gene (Abdalla & Ali, 2021) from the 25 other samples. The SuperScript IV One-Step RT-PCR System (Thermo Fisher Scientific) was used as instructed by the manufacturer. The following cycling conditions were used for the different fragments: initial reverse transcription at 45°C for 15 min, RT-deactivation/initial denaturation at 98°C for 2 min; followed by 35 cycles at 98°C for 20 s, 50°C for 30 s, 72°C for 30 s per 1 kb product length; and a final extension period of 5 min at 72°C. The amplified fragments were sent for direct sequencing by the GenSeq commercial sequencing company. The resulting sequences were deposited in GenBank under accession numbers OP585149–OP585152 (complete WMV-coding region), OP585141–OP585148, OM994910–OM994916, OM994918–OM994926 and OP821247–OP821248 (CP regions only).

The virome of the plant infected with isolate WMV-Smecno was analysed using high-throughput sequencing. Ribosomal RNA was removed from the total RNA previously isolated using a RiboMinus Plant Kit for RNA-Seq (Thermo Fisher Scientific). A NEBNext Ultra II RNA Library Prep Kit (New England Biolabs) was used to prepare

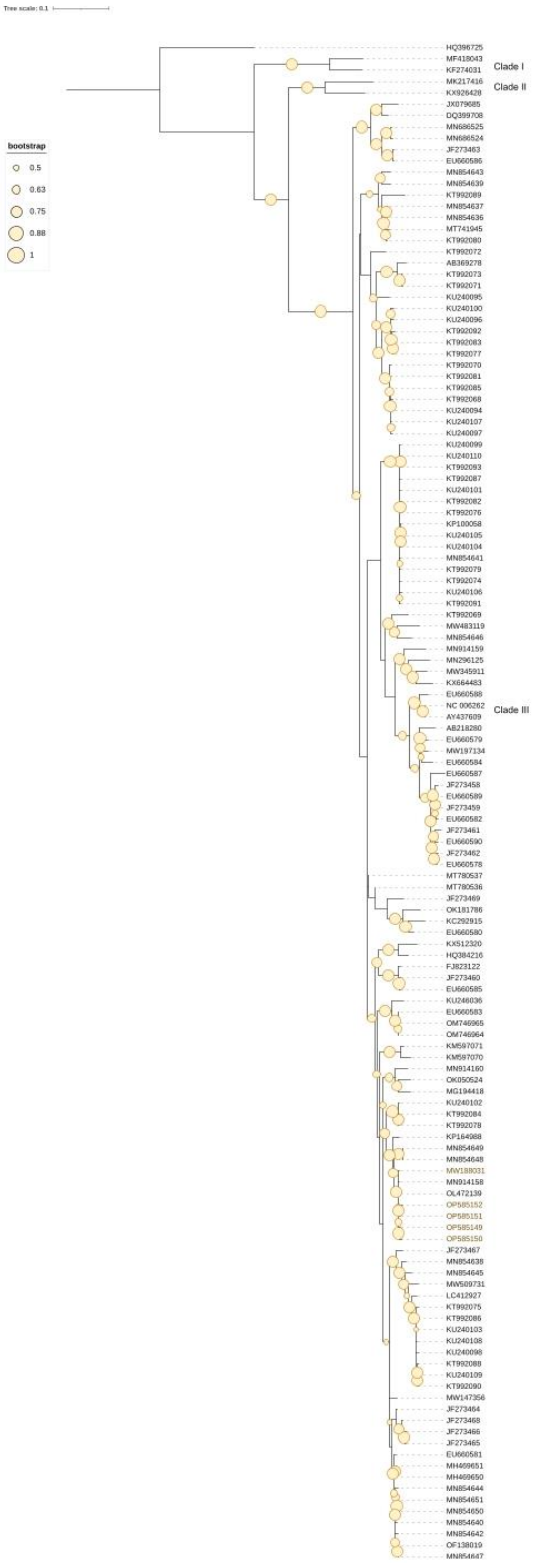

FIGURE 1 Midpoint-rooted maximum-likelihood (ML) phylogenetic tree of 133 complete coding region sequences of watermelon mosaic virus. Soybean mosaic virus (HQ396725) was set as the outgroup. Bootstrap values were set to 1000 replicates. ML phylogenetic tree was generated using MEGA X with general time reversible (GTR + G + I). Scale is 0.1 substitution/site. The five Czech isolates are in brown. The ML phylogenetic tree was viewed in iTOL. [Colour figure can be viewed at wileyonlinelibrary.com]

the libraries as instructed by the manufacturer. The latter was sequenced on a NovaSeq6000 instrument (Illumina) for 2×150 nucleotide (nt), resulting in 25,129,268 unique reads. As previously described (Komínek et al., 2019), all bioinformatics analyses, including genome assembly, were done using Geneious Prime v. 2020.2.4 software. The resulting sequence was deposited in GenBank under accession number MW188031.

2.2 | Sequence and recombination analyses

Sequences (Table S2) were aligned using the MAFFT online server (Katoh et al., 2018), and both untranslated terminal regions were trimmed to obtain the full-length coding region; for the CP coding region, the sequences were trimmed to a final fragment of 849 nt using BioEdit v. 7.2.5 (Alzohairy, 2011). For all phylogenetic analyses, the best evolutionary model was used to construct ML phylogenetic trees, and the phylogeny was tested with 1000 bootstrap replicates using MEGA X software (Kumar et al., 2018). SDT v. 1.2 was used to further confirm the phylogenetic analysis of these sequences based on pairwise sequence alignment and identity (Muhire et al., 2014). The full genomic sequences of 128 WMV isolates were retrieved from GenBank and used for phylogenetic analysis together with our five Czech WMV isolates (MW188031; OP585149–OP585152). One outgroup isolate (SbMV, HQ396725) was used to root the phylogenetic tree. The general time-reversible nucleotide substitution model with gamma distribution and invariant sites (GTR + G + I) was used to construct an ML phylogenetic tree (Figure 1). The CP region of the aforementioned sequences plus 26 Czech CP sequences (CP sequences with accession numbers OM994909–OM994926, OP585141–OP585148, OP821247–OP821248) were used to construct an ML phylogenetic tree (Figure 2) using the model with the lowest Akaike information criterion (AIC), which is a general time-reversible nucleotide substitution model with gamma distribution and invariant sites (GTR + G + I). For better visualization of any possible phylogenetic anomalies, sequences of the complete coding regions were further analysed using the SplitTree v. 4 program (Huson & Bryant, 2006). Sequences were then tested for the presence of phylogenetic anomalies using the full suite of options in RDP5 with default parameters (Boni et al., 2007; Gibbs et al., 2000; Holmes et al., 1999; Lemey et al., 2009; Martin et al., 2005, 2015; Martin & Rybicki, 2000; McGuire & Wright, 2000; Padidam et al., 1999; Posada & Crandall, 2001; Smith, 1992). Anomalies found by fewer

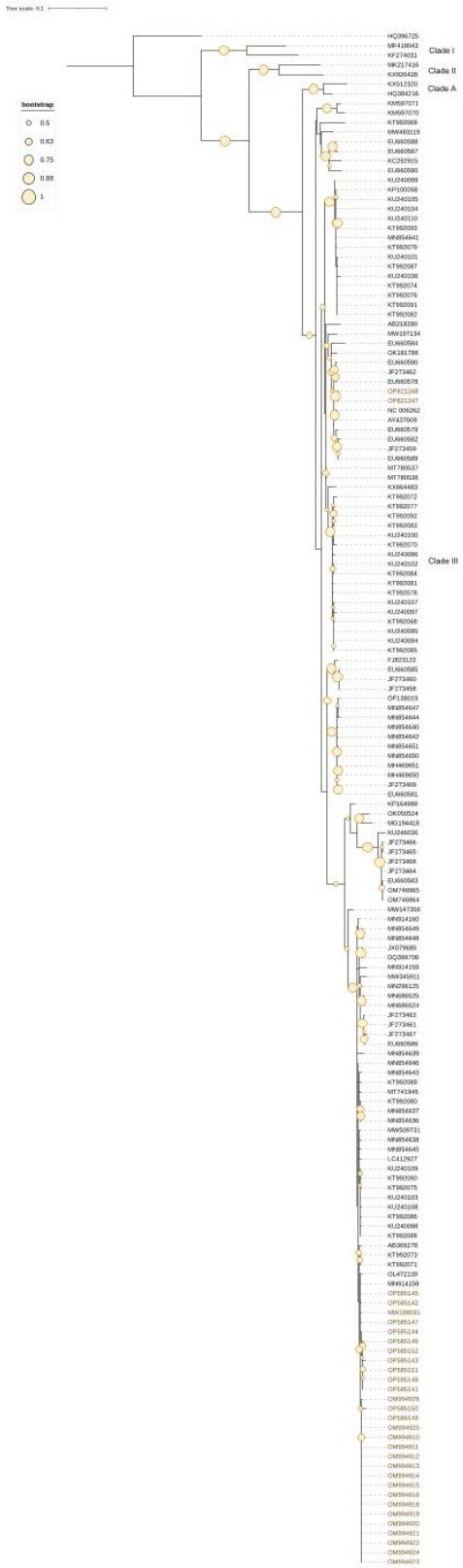

FIGURE 2 Midpoint-rooted maximum-likelihood (ML) phylogenetic tree of 159 coat protein (CP) gene sequences of watermelon mosaic virus. Soybean mosaic virus (HQ396725) was used as the outgroup. Bootstrap values were set to 1000 replicates. ML phylogenetic tree was generated using MEGA X with general time reversible (GTR + G + I). Scale is 0.1 substitution/site. The 31 Czech isolates are in brown. The ML phylogenetic tree was viewed in iTOL. [Colour figure can be viewed at wileyonlinelibrary.com]

than three methods and with $>10^{-5}$ random probability were ignored.

2.3 | Genetic variation, differentiation and gene flow analyses of WMV populations

The aligned multiple nucleotide sequences of the CP coding region were analysed using DnaSP v. 6.12 (Librado & Rozas, 2009) to determine the following: number of segregation sites (*S*), genetic diversity (π), haplotype diversity (H_d), average of nucleotide diversity (*k*) and total number of mutations (η). The neutral selection of CP among the WMV populations was examined using three statistical tests included in the DnaSP v. 6.12 program: Tajima's *D*, Fu and Li's *D*, and *F* (Librado & Rozas, 2009). The genetic differentiation and gene flow assessment between WMV populations were calculated using the fixation index F_{ST} and the effective rate of migration of gene copies into a population, N_m (*N* is effective population size, m is migration rate). The F_{ST} value ranged from 0 to 1 and showed a moderate degree of differentiation ($F_{ST} < 0.25$) and complete genetic differentiation for values above 0.25 (Hudson, 2000). The results of F_{ST} were further confirmed using three permutation-based tests: $K_S{}^*$, Z^* and $F_{ST}{}^*$ (Gao et al., 2017; Hudson et al., 1992; Valouzi et al., 2022).

2.4 | Dating analysis

Those sequences that were dated gave no evidence of a temporal signal when checked in TempEst (Rambaut et al., 2016), probably because most were recombinant (see below). Therefore, the simple "subtrees method" (Fuentes et al., 2021; Mohammadi et al., 2018) was used to extrapolate from the well-supported date estimated for PVY. A single ML phylogeny (Figure 3) describes the evolutionary distances between all its sequences using a single statistical model; thus, distances in different parts of a single ML tree can, with caution, be directly compared. The full-length genomic ORFs used by Fuentes et al. (2021) to calculate the approximate date of PVA, together with four WMV-basal sequences (Figures 1 and 2) and six reference ORFs of different molecular groups (G1, G2), subgroups (EM1–EM4) of WMV, and other BCMV-lineage viruses were downloaded from GenBank and aligned using BioEdit (Hall, 1999), MAFFT (Katoh et al., 2018) and the PAL2NAL online server (http://www.bork.embl.de/pal2nal/, Suyama et al., 2006). ML phylogenetic trees were calculated using PhyML 3.0 (Guindon & Gascuel, 2003), and

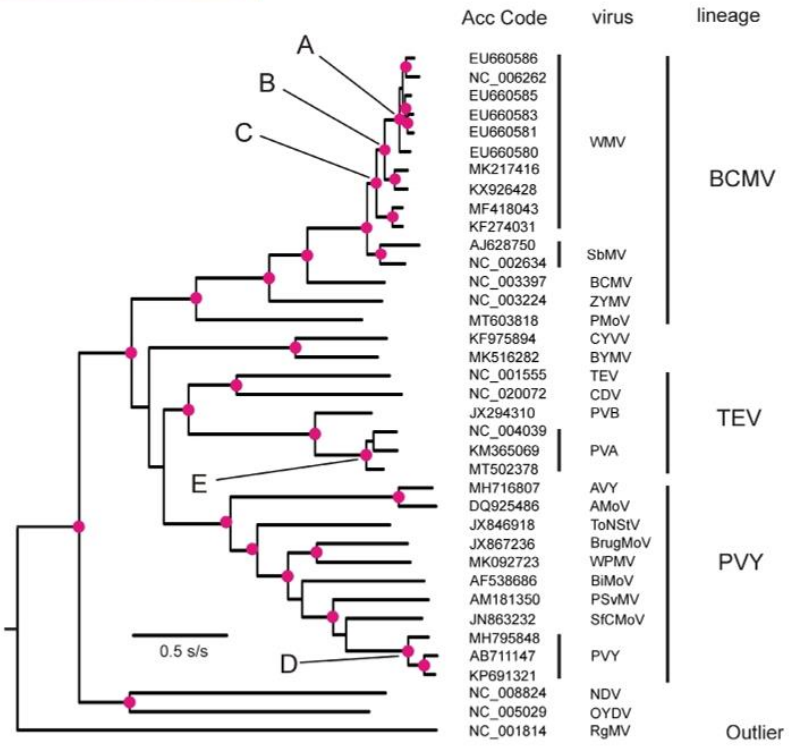

FIGURE 3 Maximum-likelihood (ML) phylogeny of a representative selection of watermelon mosaic virus (WMV) genomic open reading frames together with those used by Fuentes et al. (2019, 2021) for dating the phylogenies of potato viruses Y and A. The mean evolutionary distance of node A was 0.138 ± 0.015 substitutions/site (s/s); node B 0.282 ± 0.015 s/s; node C 0.334 ± 0.019 s/s; node D 0.256 ± 0.008 s/s. The virus acronyms are AMoV, arracacha mottle virus; AVY, arracacha virus Y; BCMV, bean common mosaic virus; BiMoV, bidens mottle virus; BrugMoV, Brugmansia mosaic virus; BYMV, bean yellow mosaic virus; CDV, Colombian datura virus; CYVV, clover yellow vein virus; NDV, narcissus degeneration virus; OYDV, onion yellow dwarf virus; PMoV, peanut mottle virus; PSvMV, pepper severe mosaic virus; PVA, potato virus A; PVB, potato virus B; PVY, potato virus Y; RgMV, ryegrass mosaic rymovirus; SbMV, soybean mosaic virus; SfCMoV, sunflower chlorotic mottle virus; TEV, tobacco etch virus; ToNStV, tomato necrotic stunt virus; WMV, watermelon mosaic virus; WPMV, wild potato mosaic virus; and ZYMV, zucchini yellow mosaic virus. [Colour figure can be viewed at wileyonlinelibrary.com]

the SH method was used to assess the statistical support of their topologies (Shimodaira & Hasegawa, 1999). Trees were visualized using FigTree v. 1.3 (http://tree.bio.ed.ac.uk/software/figtree/) and a commercial graphics package. PATRISTIC (Fourment & Gibbs, 2006) was used to convert complete tree files to matrices of all pairwise patristic distances between the tips. The positions of the basal nodes of the selected subtrees in ML phylogenies were estimated as the mean pairwise patristic distance between all tips connected through those nodes (Hajizadeh et al., 2019; Mohammadi et al., 2018) using Excel (Microsoft).

3 | RESULTS

3.1 | Sequence and recombination analysis

The genomic sequence of the WMV-Smecno isolate was obtained as described above. A de novo contig from high-throughput sequencing with a length of 9896 nt was identified as WMV and was used as a reference for reads mapping. The total number of mapped WMV reads was 489,574, covering 100% of the total viral genome. No other viruses were detected in the sequencing data. The genome length of WMV was found to be 10,027 nt. It presents a large ORF starting at position 128 and ending at position 9775. Another four full genomes were amplified by RT-PCR using a set of primers and assembled into a final product of 9642 nt, presenting the coding region of WMV that encodes a large polyprotein with 3215 amino acids. Twenty-six CP coding regions of WMV were amplified to generate fragments of 849 nt (283 amino acids).

RDP5 (v. 5.34) was used for recombination analysis of 133 complete genomes (including 128 from GenBank and the five Czech WMV isolates from this study) and revealed 80 recombination events, of which 63 were located in two-thirds of the coding region (Table S3). The RDP5 results showed that only five WMV isolates were nonrecombinant (KX926428, MF418043, MK217416, EU660580 and KC292915), and three of these nonrecombinant isolates formed basal clusters of WMV, which we called Clades I and II (Figures 1 and 2); all the other WMV isolates were in Clade III. The

frequency of recombinant WMV isolates was 96.24%. Most putative parents of the detected recombination events were of Asian origin. Intraspecific recombination was not limited to a single breakpoint, but to several locations. The frequency of those with a single intraspecific recombination was approximately 36% (Table S4), and all the others were multiple (61%; Table S5). The highest recombination rate (28%) was observed for the coding region involving parts of the P1 gene (P1 or the combined region of P1-HCPro and P1-CI coding regions). The reference isolates previously described by Desbiez et al. (2011) (CL: AY437609; G2: EU660580; EM1: EU660581; EM2: EU660583; EM3: EU660586; EM4: EU660585), used to determine the different molecular groups, were considered to be nonrecombinant. However, the current RDP5 analysis of a larger number of sequences demonstrates that these reference sequences, except for EU660580, are in fact recombinants. The five Czech WMV isolates, the subject of our study, had a 4385 nt recombinant fragment in the CI-CP coding region. These sequences have two postulated "parents" (i.e., the sequences closest to those contributing the different parts of the recombinant sequences): the larger fragment is closest to KT992093, obtained from the non-cucurbit host *Panax ginseng* originating in South Korea, and the minor "parent" is closest to EU660586 isolated in France. One Czech isolate (OP585152) contained an extra recombinant fragment of 200 nt in the 6K1-CI coding region. The putative parents detected for this event were OP585150 as the major parent and the minor parent was MN854642, originating from the Czech Republic and South Korea, respectively.

3.2 | Phylogenetics and identity matrix

The phylogenetic tree of the complete sequences (Figure 1) shows three well-supported clades. Clade I contained two isolates, MF418043 and KF27403, Clade II also contained two isolates KX926428 and MK217416. The remaining isolates clustered within Clade III. The isolates clustered within Clades I and II were of Asian origin, specifically Chinese, and were extracted from non-cucurbitaceous hosts (*Alcea rosea*, *Ailanthus altissima* and *P. ginseng*). Clade III was the largest molecular group, and its members had no consistency of host or geographical origin. The topology of the tree was further confirmed using SDT v. 1.2 software to calculate the pairwise genetic identities between the complete genome sequences of these 133 WMV isolates (Figure S1). The members of Clades I and II differed in pairwise genetic identities from Clade III isolates by 81% to 95%. The five Czech isolates clustered within Clade III, close to two isolates (99.1% to 99.6%): OL472139, which was extracted from pools of unidentified weeds from Slovenia, and MN914158, which was extracted from squash originating from France (Figure S1).

The ML tree constructed based on the CP gene (Figure 2) showed two basal clades containing the same isolates as the phylogeny of the complete sequences. It also had an extra basal clade, Clade A, containing sequences from two isolates, KX512320 and HQ384216, isolated from non-cucurbitaceous hosts and originating from the United States; these are discussed below. Many of the groupings in the phylogenies from sequences in the CP and the full coding region were not the same, confirming that they were recombinants. This was also confirmed by reticulation of the SplitsTree networks (Figure 4), which indicated the presence of significant recombination throughout the phylogeny except for Clades I and II. Based on the SDT v. 1.2 program, the sequence similarity among the Czech samples was 95.4%–100% and 93.8%–100% at the nucleotide and amino acid levels, respectively. All Czech isolates clustered within Clade III, in which two of them (OP821247 and OP821248) were closely related to AY437609, which originated from France and were isolated from pumpkin with pairwise sequence identity ranging between (98.8%–99.6%) and (99.3%–98.9%) at nucleotide and amino acid levels, respectively, whereas all the remaining 29 isolates shared the highest sequence identity with OL472139 (98.8%–99.6% nt), and MN914158 (99.2%–100% nt), and their sequence identity based on the amino acid sequence with these two isolates ranged from 99.3% to100% (Figures S2 and S3).

3.3 | Genetic variation, differentiation and gene flow analyses of WMV populations

The DNaSP v.0.6.12 program was used to estimate the genetic diversity of the CP region (Table S6). The Czech population had the smallest genetic diversity, $\pi = 0.010 \pm 0.00013$, with an average nucleotide diversity number $k = 9.07$. They also had a small number of polymorphic sites ($S = 65$) and number of mutations ($\eta = 67$); the number of haplotypes (H) was 15, and haplotype diversity (H_d) was 0.76. The highest genetic diversity was observed in the Chinese population, $\pi = 0.11 \pm 0.001$. The three statistical tests, Tajima's *D*, Fu and Li's *D* and Fu and Li's *F*, of the Czech population were negative with none of them statistically significant. The populations of the newly observed phylogroups I, II and A were demonstrated to have a high genetic diversity $\pi = 0.10 \pm 0.002$, 0.09 ± 0.02 and 0.04 ± 0.011, respectively. All populations in the different phylogenetic groups were genetically distinct because they had a high fixation index value ($F_{ST} > 0.5$) and all statistical tests were significant.

Infrequent gene flow was observed between the Czech population and other populations from different geographical regions ($F_{ST} > 0.25$). All the populations examined were genetically differentiated from each other, and these results were confirmed by significant F_{ST}^*, Z^* and S_{nn} independent tests (Table S7).

3.4 | Molecular dating

It was not possible to estimate the dates of nodes in the WMV phylogeny by tip dating because the dated sequences gave no evidence of a temporal signal using TempEst (Rambaut et al., 2016), and most WMV genomes were recombinants. Therefore, we used the strategy of subtree dating, a method used by Mohammadi et al. (2018) and Fuentes et al. (2021). This method involves estimating and comparing the positions of the nodes in a single ML phylogeny

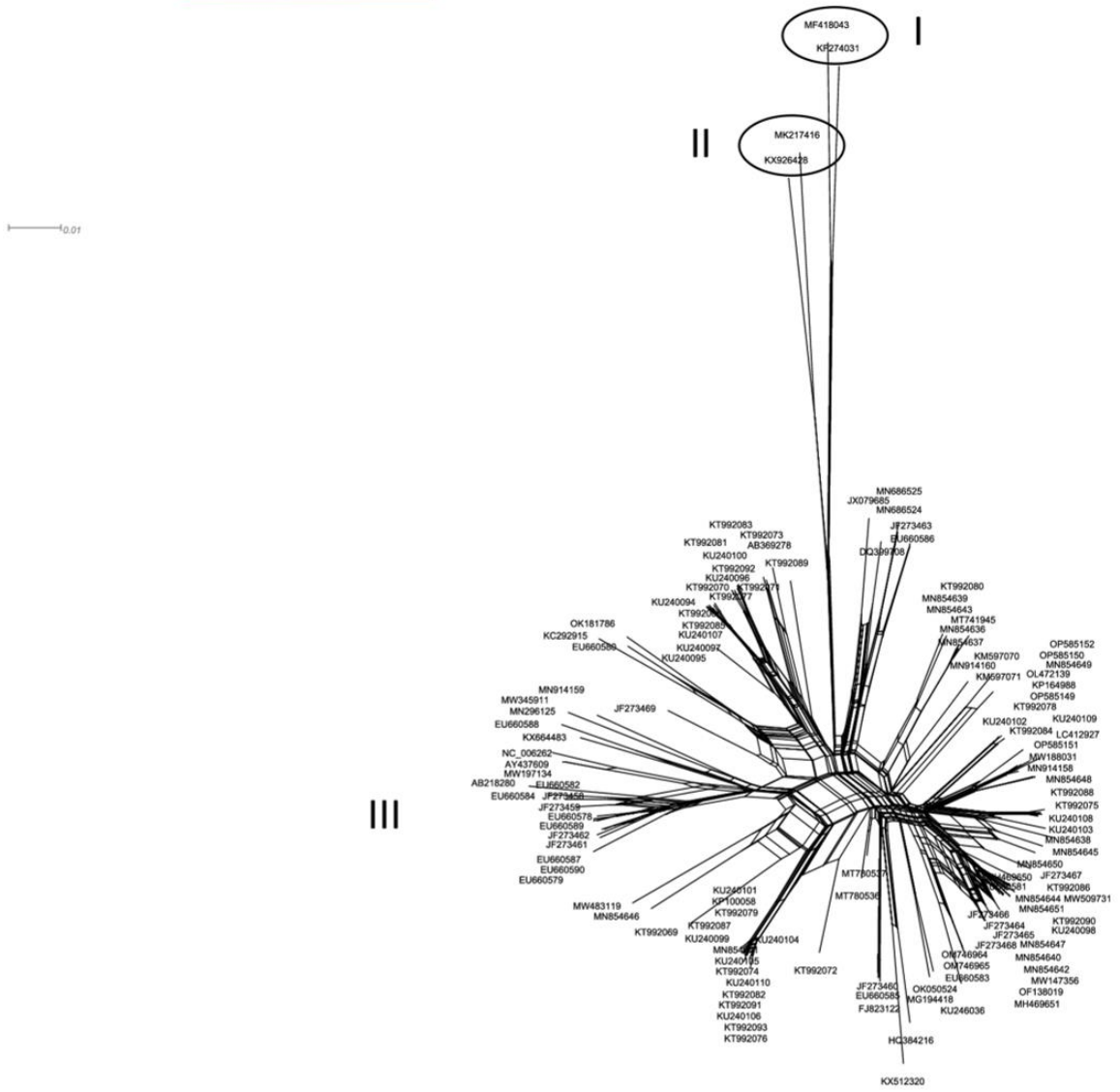

FIGURE 4 Phylogenetic networks examining 133 complete coding regions of watermelon mosaic virus and showing the three clades (I, II, III); created by SplitsTree v. 4.17.2.

from their mean pairwise patristic distances, and their dates calculated by simple arithmetic using the date of a well-established "time to the most recent common ancestor" (TMRCA) of a related reference virus included in the phylogeny. The phylogeny (Figure 3) used for dating WMV was calculated from sequences representing both basal branches of all nodes to be dated. Therefore, it included both sequences of Clade I (KF274031 and MF418043), which were isolated in China from *A. altissima* (Ding et al., 2006); and both of Clade II (MK217416, a Chinese isolate from *A. rosea*; and KX926428, a Chinese isolate from *P. ginseng*; Park et al., 2017). It also contained all the sequences used by Fuentes et al. (2019, 2021) to compare the nodal dates of PVY and PVA. The basal node of the PVY phylogeny was estimated by Fuentes et al. (2019) to be 156 CE and by Gao et al. (2020) to be 158 CE. The mean pairwise patristic distance of node D was 0.256794 ± 0.008103, and that of node B was 0.282347 ± 0.015101. Fuentes et al. (2021) used a TMRCA of PVY = 157 CE as a mean value retrieved from both publications of Fuentes and Gao and their co-authors in 2019 and 2020. Therefore, node A is estimated to be dated 1009.4 years before present (YBP) or 1014 CE; node B to be 2027.4 YBP or 4.4 BCE; and node C 2433.8 YBP or 411 BCE. The mean coefficients of variation are 0.059, 0.115 and 0.053, respectively.

4 | DISCUSSION

WMV in ornamental pumpkin was first detected in the Czech Republic in 2011 by Svoboda using serological tests. In the present investigation, we used gene sequencing to confirm and extend the earlier study, and to check whether gene sequences would provide more detailed information of the immediate origins of the Czech WMV population and also estimate the historical origin of the entire WMV population.

ML phylogenetic trees were constructed based on the sequences of both the complete coding region and the CP gene; these trees showed a different topology from those previously reported (Desbiez et al., 2011), but, like that of Wang et al. (2017), identified a sequence from a Chinese isolate as basal to the other WMV sequences. Here, we report that the NCBI database now contains three additional sequences that group with KF274031 forming two distinct basal Clades, I and II, with all others forming Clade III. Clade I and II contained four isolates, all of which originated from China, and were isolated from non-cucurbit hosts (Clade I; MF418043 and KF274031, and Clade II; MK217416 and KX926428); Clade III includes all the other 129 sequences from isolates from different geographical locations (Asia, Europe and the United States), and from different plant hosts (cucurbitaceous and non-cucurbitaceous). Sequences of Clades I and II have identities with those of Clade III that range from 81.1% to 93.0% and 85.1% to 96.8% at nucleotide and amino acid levels of the CP, respectively.

An RDP analysis revealed that three of the four sequences of Clades I and II were nonrecombinant; that of KF274031 had a small recombinant region with a WMV "parent". In contrast, almost all other genomic WMV sequences were recombinant (96.2%); this is the largest percentage of recombinants in the population of known potyviruses, even more than those in turnip mosaic virus populations, which are three-quarters recombinant (Yasaka et al., 2017). RDP5 analysis also detected one statistically significant recombination event in the five Czech isolates; their closest isolates were from France, and from the non-cucurbit host, *P. ginseng*, collected in South Korea.

The RDP5 analysis also showed that recombination was not limited to a single breakpoint, but occurred throughout the genome with no clear pattern, despite earlier reports with fewer sequences (Bertin et al., 2020; Desbiez & Lecoq, 2008; Moreno et al., 2004; Perrot et al., 2021). The large number of recombination breakpoints in the P1 coding region could explain the broad host range of WMV, as P1 is believed to be involved in adaptation to new hosts and symptom determinants (Maliogka et al., 2012); the large number of recombination events in the regions involving the CP coding region could be linked to the key role played by this gene in the transmissibility of potyviruses by aphids (Gadhave et al., 2020). Recombination has been linked to emergence of new viral species, increased pathogenicity, broad host range and the ability to break resistance (LaTourrette & Garcia-Ruiz, 2022).

The CP tree has the same two basal clades, I and II, as the complete ORF tree and both trees agree that there are no major recombinant regions in them. However, the CP tree has a third basal clade (Clade A in Figure 2) of two isolates (KX512320 and HQ384216). These CPs are of interest as they resemble Clade I and II isolates in having hosts (*Passiflora edulis* and *Dendrobium anosmum*) that are not watermelons, but, unlike the Clade I and II isolates, were collected in the United States instead of China. Thus, it is possible that these represent another basal clade of WMV; however, the other regions of the sequences of these two isolates provide no useful information, as they are triple recombinants.

Desbiez et al. (2011) placed the sequences of the WMV population known at that time into several molecular groups and subgroups, all of which we now call Clade III; however, it is clear that these groupings within Clade III will mostly reflect the phylogenetic dominance of difference regions of the genome and the recombinant regions in them. Some Czech isolates were taken from the collection of VURV (CRI Prague) and were isolated 10 years ago, whereas the others were collected recently ($n=29$). Nonetheless, there were 15 haplotypes among the 31 Czech isolates, which had the lowest haplotype diversity ($H_d=0.76$). Furthermore, the Czech WMV population had the lowest genetic diversity ($\pi=0.010\pm0.00013$) compared to the other populations tested, indicating less evolutionary divergence and recombination compared with the Chinese population and those from South Korea, France, Italy and the United States. All populations of different phylogenetic groups were shown to be genetically differentiated with no detectable gene flow between them, as judged by their CP genes ($F_{ST}>0.25$, $N_m<1$), and with statistically significant K_{ST}^*, Z^* and S_{nn} independent tests.

The members of Clades I and II had a higher estimated genetic diversity ($\pi=0.14\pm0.006$) than members of Clade III ($\pi=0.04\pm0.00001$). In addition, the members of Clades I and II had the longest branch lengths, confirming that the two basal clades are older than those of Clade III. The date of these nodes was estimated (Figure 3) by the simple subtree method using a single ML phylogeny and found nodes B and C (at which the basal sequences diverge) diverged around 2434 and 2027 YBP, respectively, and the basal node of clade III at 1009 YBP. Watermelon was domesticated in East Africa 4000–6000 years ago and its use spread throughout the Roman Empire. There is some discussion about when watermelon was first grown in China. Dane and Liu (2007) stated that watermelon was introduced to India in the 9th century, and to China by the 11th century; Simoons (2014) states that watermelon arrived in China "only in the twelfth century"; and the most complete account is in *Life of Guang Zhou* (Anonymous, 2021), which explains that although Spengler (2019) stated that watermelon may have arrived in south China between 200 BCE and 200 CE via the maritime Silk Road, it was not grown widely until after being taken via the overland Silk Road through Mongolia to Liao Shangjing in northern China during the Five Dynasties (907–960 CE).

Thus, the simplest historical scenario is that WMV originated as a virus infecting various plant species (*A. altissima*, *A.a rosea*, *P.*

ginseng, etc.) in north China at least 2000 years ago during the Zhou and Han dynasties (411 BC and 4.4 BC), but moved from non-cucurbit hosts to watermelon around 1000 years ago when watermelon was first grown as a crop in northern China during the Five Dynasties (907–960 CE). Furthermore, the dating and origin of WMV estimated in this study are in agreement with an earlier study suggesting that WMV is a member of the BCMV lineage that emerged more than 3000 years ago and dispersed from South-east and East Asia (Gibbs et al., 2008, 2020).

This study demonstrated that WMV is a highly recombinogenic potyvirus that has been dispersed worldwide through the movement of infected plant materials and, locally, by vectors (Gagarinova et al., 2008). WMV probably emerged and became a significant pathogen of the watermelon crop in north China approximately 1000 years ago, raising the question of whether the virus was pre-adapted to become a significant pathogen of a novel crop, or whether specific adaptations occurred when it invaded the crop. WMV adaptation to trees is further supported by Laney et al. (2012) who reported that both WMV and papaya ringspot virus infected *Robinia pseudoacacia* in north-west Arkansas in the United States and were significantly seedborne (48% and 42%) in this species. This indicates that it would be interesting to compare the rates of seed transmission in watermelon of WMV isolates of the different clades.

ACKNOWLEDGEMENTS

The research performed at CRI, Prague was supported by the Ministry of Agriculture of the Czech Republic, institutional support MZE-RO0423.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon request.

ORCID

Karima Ben Mansour https://orcid.org/0000-0002-0054-5954
Petr Komínek https://orcid.org/0000-0001-7651-5236

REFERENCES

Abdalla, O.A. & Ali, A. (2021) Genetic variability and evidence of a new subgroup in watermelon mosaic virus isolates. *Pathogens*, 10, 1245.

Alzohairy, A.M. (2011) BioEdit: an important software for molecular biology. *GERF Bulletin of Biosciences*, 2, 60–61.

Anonymous. (2021) *The introduction of watermelon into China*. Available from: https://www.lifeofguangzhou.com/wap/ [Accessed 30th March 2023]

Bertin, S., Mangili, A., McLeish, M. & Tomassoli, L. (2020) Genetic variability of watermelon mosaic virus isolates infecting cucurbit crops in Italy. *Archives of Virology*, 165, 937–946.

Boni, M.F., Posada, D. & Feldman, M.W. (2007) An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics*, 176, 1035–1047.

Chare, E.R. & Holmes, E.C. (2006) A phylogenetic survey of recombination frequency in plant RNA viruses. *Archives of Virology*, 151, 933–946.

Chung, B.Y.W., Miller, W.A., Atkins, J.F. & Firth, A.E. (2008) An overlapping essential gene in the *Potyviridae*. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 5897–5902.

Cohen, S. & Nitzany, F.E. (1963) Identity of viruses affecting cucurbits in Israel. *Phytopathology*, 53, 193–196.

Dane, F. & Liu, J. (2007) Diversity and origin of cultivated and citron type watermelon (*Citrullus lanatus*). *Genetic Resources and Crop Evolution*, 54, 1255–1265.

Desbiez, C., Costa, C., Wipf-Scheibel, C., Girard, M. & Lecoq, H. (2007) Serological and molecular variability of watermelon mosaic virus (genus *Potyvirus*). *Archives of Virology*, 152, 775–781.

Desbiez, C., Joannon, B., Wipf-Scheibel, C., Chandeysson, C. & Lecoq, H. (2009) Emergence of new strains of *Watermelon mosaic virus* in south-eastern France: evidence for limited spread but rapid local population shift. *Virus Research*, 141, 201–208.

Desbiez, C., Joannon, B., Wipf-Scheibel, C., Chandeysson, C. & Lecoq, H. (2011) Recombination in natural populations of watermelon mosaic virus: new agronomic threat or damp squib? *Journal of General Virology*, 92, 1939–1948.

Desbiez, C. & Lecoq, H. (2004) The nucleotide sequence of watermelon mosaic virus (WMV, *Potyvirus*) reveals interspecific recombination between two related potyviruses in the 5′ part of the genome. *Archives of Virology*, 149, 1619–1632.

Desbiez, C. & Lecoq, H. (2008) Evidence for multiple intraspecific recombinants in natural populations of watermelon mosaic virus (WMV, *Potyvirus*). *Archives of Virology*, 153, 1749–1754.

Ding, J., Wu, Y., Zheng, H., Fu, W., Reardon, R. & Liu, M. (2006) Assessing potential biological control of the invasive plant, tree-of-heaven, *Ailanthus altissima*. *Biocontrol Science and Technology*, 16, 547–566.

Fourment, M. & Gibbs, M.J. (2006) PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. *BMC Evolutionary Biology*, 6, 1.

Fuentes, S., Gibbs, A.J., Adams, I.P., Wilson, C., Botermans, M., Fox, A. et al. (2021) Potato virus A isolates from three continents: their biological properties, phylogenetics, and prehistory. *Phytopathology*, 111, 217–226.

Fuentes, S., Jones, R.A.C., Matsuoka, H., Ohshima, K., Kreuze, J. & Gibbs, A.J. (2019) Potato virus Y; the Andean connection. *Virus Evolution*, 5, vez037.

Gadhave, K.R., Gautam, S., Rasmussen, D.A. & Srinivasan, R. (2020) Aphid transmission of *Potyvirus*: the largest plant-infecting RNA virus genus. *Viruses*, 12, 773.

Gagarinova, A.G., Babu, M., Strömvik, M.V. & Wang, A. (2008) Recombination analysis of *Soybean mosaic virus* sequences reveals evidence of RNA recombination between distinct pathotypes. *Virology Journal*, 5, 1–8.

Gao, F., Kawakubo, S., Ho, S.Y.W. & Ohshima, K. (2020) The evolutionary history and global spatio-temporal dynamics of potato virus Y. *Virus Evolution*, 6, veaa056.

Gao, F., Zou, W., Xie, L. & Zhan, J. (2017) Adaptive evolution and demographic history contribute to the divergent population genetic structure of *Potato virus Y* between China and Japan. *Evolutionary Applications*, 10, 379–390.

Gao, Q., Ren, H.L., Xiao, W.Y., Zhang, Y., Zhou, B. & Xu, D.L. (2021) First report of watermelon mosaic virus causing a mosaic disease on *Cucumis metuliferus* in China. *Plant Disease*, 105, 2025.

García-Arenal, F., Fraile, A. & Malpica, J.M. (2001) Variabilty and genetic structure of plant virus populations. *Annual Review of Phytopathology*, 39, 157–186.

Gibbs, A.J., Hajizadeh, M., Ohshima, K. & Jones, R.A.C. (2020) The potyviruses: an evolutionary synthesis is emerging. *Viruses*, 12, 132.
Gibbs, A.J., Ohshima, K., Yasaka, R., Mohammadi, M., Gibbs, M.J. & Jones, R.A.C. (2017) The phylogenetics of the global population of potato virus Y and its necrogenic recombinants. *Virus Evolution*, 3, vex002.
Gibbs, A.J., Trueman, J.W.H. & Gibbs, M.J. (2008) The bean common mosaic virus lineage of potyviruses: where did it arise and when? *Archives of Virology*, 153, 2177–2187.
Gibbs, M.J., Armstrong, J.S. & Gibbs, A.J. (2000) Sister-scanning: a Monte Carlo procedure for assessing signals in rebombinant sequences. *Bioinformatics*, 16, 573–582.
Guindon, S. & Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52, 696–704.
Hajizadeh, M., Gibbs, A.J., Amirnia, F. & Glasa, M. (2019) The global phylogeny of *Plum pox virus* is emerging. *Journal of General Virology*, 100, 1457–1468.
Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95–98.
Holmes, E.C., Worobey, M. & Rambaut, A. (1999) Phylogenetic evidence for recombination in dengue virus. *Molecular Biology and Evolution*, 16, 405–409.
Hudson, R.R. (2000) A new statistic for detecting genetic differentiation. *Genetics*, 155, 2011–2014.
Hudson, R.R., Boos, D.D. & Kaplan, N.L. (1992) A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution*, 9, 138–151.
Huson, D.H. & Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23, 254–267.
Inoue-Nagata, A.K., Jordan, R., Kreuze, J., Li, F., López-Moya, J.J., Mäkinen, K. et al. (2022) ICTV virus taxonomy profile: *Potyviridae* 2022. *Journal of General Virology*, 103, 1–2.
Katoh, K., Rozewicki, J. & Yamada, K.D. (2018) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20, 1160–1166.
Komínek, P., Massart, S., Pham, K., van Leeuwen, P. & Komínková, M. (2019) Characterisation of a novel virus infecting orchids of the genus *Pleione*. *Virus Research*, 261, 56–59.
Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35, 1547–1549.
Laney, A.G., Avanzato, M.V. & Tzanetakis, I.E. (2012) High incidence of seed transmission of *Papaya ringspot virus* and *Watermelon mosaic virus*, two viruses newly identified in *Robinia pseudoacacia*. *European Journal of Plant Pathology*, 134, 227–230.
LaTourrette, K. & Garcia-Ruiz, H. (2022) Determinants of virus variation, evolution, and host adaptation. *Pathogens*, 11, 1039.
Lemey, P., Lott, M., Martin, D.P. & Moulton, V. (2009) Identifying recombinants in human and primate immunodeficiency virus sequence alignments using quartet scanning. *BMC Bioinformatics*, 10, 126.
Librado, P. & Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, 1451–1452.
Maliogka, V.I., Salvador, B., Carbonell, A., Sáenz, P., León, D.S., Oliveros, J.C. et al. (2012) Virus variants with differences in the P1 protein coexist in a plum pox virus population and display particular host-dependent pathogenicity features. *Molecular Plant Pathology*, 13, 877–886.
Martin, D.P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. (2015) RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, 1, vev003.
Martin, D.P., Posada, D., Crandall, K.A. & Williamson, C. (2005) A modified Bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Research and Human Retroviruses*, 21, 98–102.
Martin, D.P. & Rybicki, E. (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics*, 16, 562–563.
McGuire, G. & Wright, F. (2000) TOPAL 2.0: improved detection of mosaic sequences within multiple alignments. *Bioinformatics*, 16, 130–134.
Mohammadi, M., Gibbs, A.J., Hosseini, A. & Hosseini, S. (2018) An Iranian genomic sequence of *Beet mosaic virus* provides insights into diversity and evolution of the world population. *Virus Genes*, 54, 272–279.
Moreno, I.M., Malpica, J.M., Díaz-Pendón, J.A., Moriones, E., Fraile, A. & García-Arenal, F. (2004) Variability and genetic structure of the population of watermelon mosaic virus infecting melon in Spain. *Virology*, 318, 451–460.
Muhire, B.M., Varsani, A. & Martin, D.P. (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One*, 9, e108277.
Padidam, M., Sawyer, S. & Fauquet, C.M. (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology*, 265, 218–225.
Park, D., Kim, H. & Hahn, Y. (2017) Genome sequence of a distinct watermelon mosaic virus identified from ginseng (*Panax ginseng*) transcriptome. *Acta Virologica*, 61, 479–482.
Perrot, S., Houdault, S., Grimault, V., Lecoq, H., Buisson, M., Bertrand, F. et al. (2021) VIRUS COURGETTE – amélioration de l'évaluation de la résistance de la courgette aux *Cucumber mosaic virus* (CMV), *Zucchini yellow mosaic virus* (ZYMV) et *Watermelon mosaic virus* (WMV). *Innovations Agronomiques*, 84, 33–43.
Posada, D. & Crandall, K.A. (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 13757–13762.
Rambaut, A., Lam, T.T., Carvalho, L.M. & Pybus, O.G. (2016) Exploring the temporal structure of heterochronous sequences using TempEst (formerly path-O-gen). *Virus Evolution*, 2, vew007.
Shimodaira, H. & Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, 16, 1114–1116.
Simoons, F.J. (2014) Food in China: a cultural and historical inquiry. CRC Press. Available from: https://www.perlego.com/book/1663351/food-in-china-a-cultural-and-historical-inquiry-pdf [Accessed 30th March 2023].
Smith, J.M. (1992) Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*, 34, 126–129.
Spengler, R.N. (2019) *Fruit from the sand: the Silk Route origins of the foods we eat*. Berkeley, CA: University of California Press.
Suyama, M., Torrents, D. & Bork, P. (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, 34, 609–612.
Svoboda, J. (2011) Pathogenic viruses on cucurbitaceous vegetables and their spread in the Czech Republic. *Acta Horticulturae*, 917, 309–315.
Valouzi, H., Shahmohammadi, N., Golnaraghi, A., Moosavi, M.R. & Ohshima, K. (2022) Genetic diversity and evolutionary analyses of potyviruses infecting narcissus in Iran. *Journal of Plant Pathology*, 104, 237–250.
Varsani, A., Shepherd, D.N., Monjane, A.L., Owor, B.E., Erdmann, J.B., Rybicki, E.P. et al. (2008) Recombination, decreased host specificity and increased mobility may have driven the emergence of maize streak virus as an agricultural pathogen. *Journal of General Virology*, 89, 2063–2074.
Wang, D., Li, G. & Du, S.S. (2017) Occurrence of viruses infecting melon in Xinjiang of China and molecular characterization of *Watermelon mosaic virus* isolates. *European Journal of Plant Pathology*, 147, 919–931.

Yasaka, R., Fukagawa, H., Ikematsu, M., Soda, H., Korkmaz, S., Golnaraghi, A. et al. (2017) The timescale of emergence and spread of turnip mosaic potyvirus. *Scientific Reports*, 7, 4240.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ben Mansour, K., Gibbs, A.J., Komínková, M., Komínek, P., Brožová, J., Kazda, J. et al. (2023) Watermelon mosaic virus in the Czech Republic, its recent and historical origins. *Plant Pathology*, 72, 1528–1538. Available from: https://doi.org/10.1111/ppa.13766

II. **Study 2**

-Title: Grapevine Pinot gris virus in Germany: From where did the virus come, and when?

-Brief description: Grapevine Pinot gris virus (GPGV) isolates were obtained from German vineyards with the objective of investigating their diversity. An understanding of the time and place of emergence of a plant virus and the evolution of its population is crucial for the elucidation of its biology, epidemiology and for the development of effective management strategies. Phylogenetic and dating analyses of these and GPGV genes and genomes available in GenBank demonstrated that the virus likely diverged from grapevine berry inner necrosis virus (GINV) in wild and cultivated *Vitis* species, notably *Vitis coignetiae*, which is native to North-east Asia. This divergence is estimated to have occurred approximately 3500 years ago. The German isolates exhibited greater genetic diversity than those of other European populations, particularly with regard to the MP and CP genes. This suggests that the initial stages of the GPGV invasion of Europe were in Germany, rather than in Italy.

Received: 2 September 2023 | Accepted: 14 October 2023

DOI: 10.1111/ppa.13824

Plant Pathology WILEY

ORIGINAL ARTICLE

Grapevine Pinot gris virus in Germany: From where did the virus come, and when?

Karima Ben Mansour[1,2,3] | Adrian J. Gibbs[4] | Noemi Meßmer[5] | René Fuchs[5] | Thierry Wetzel[1] | Patrick Winterhagen[1]

[1]DLR Rheinpfalz, Institute of Plant Protection, Neustadt a. d. Weinstraße, Germany

[2]Department of Plant Protection, Faculty of Agrobiology, Food, and Natural Resources, Czech University of Life Sciences Prague, Prague, Czech Republic

[3]Ecology, Diagnostics and Genetic Resources of Agriculturally Important Viruses, Fungi and Phytoplasmas, Crop Research Institute, Prague, Czech Republic

[4]Emeritus Faculty, Australian National University, Canberra, Australian Capital Territory, Australia

[5]State Institute of Viticulture and Enology, Freiburg, Germany

Correspondence
Karima Ben Mansour, Institute of Plant Protection, DLR Rheinpfalz, Neustadt a. d. Weinstraße, Germany.
Email: karina79@hotmail.fr

Abstract

Grapevine Pinot gris virus (GPGV) isolates were obtained from German vineyards to investigate their diversity. Phylogenetic and dating analyses of these and GPGV genes and genomes available in GenBank showed that the virus probably diverged from grapevine berry inner necrosis virus (GINV) in wild and cultivated *Vitis* species, notably *Vitis coignetiae*, growing in North-east Asia around 3500 years ago. GPGV probably infected the Eurasian grape (*Vitis vinifera* subsp. *vinifera*) when those cultivars were first taken to China during the Han Dynasty (226 BCE–220 CE). GPGV then spread to Europe around 1800 CE, probably via the dissemination of infected plants, and from there, eventually spread worldwide. German isolates were only found in all parts of the post-1800 CE phylogeny. The German isolates were genetically more diverse, for both MP and CP genes, than those of other European populations, suggesting that the initial stages of the GPGV invasion of Europe were in Germany, not Italy. We discuss the likely North-east Asian origin of both GPGV and GINV, and the possible coincidences of phylogenetic date estimates with changes to European and world viticulture practices.

KEYWORDS
genetic variability, grapevine Pinot gris virus, molecular dating, phylogenetic analysis, *Vitis coignetiae*

1 | INTRODUCTION

More than 90 taxonomically distinct viruses have been isolated from grapevines (Al Rwahnih et al., 2021; Fuchs, 2020; Giampetruzzi et al., 2012), but only two, grapevine Pinot gris virus (GPGV) and grapevine berry inner necrosis virus (GINV), are members of the *Trichovirus* genus of the *Betaflexiviridae* family (Fan et al., 2017; Yoshikawa et al., 1997). Like other members of the genus, they have flexuous filamentous virions around 700 nm in length, each of which contains a single-stranded, positive-sense RNA genome with three overlapping open reading frames encoding a replicase (RdRp), movement protein (MP) and coat protein (CP) (Giampetruzzi et al., 2012).

GPGV, first reported in Italy in 2012 (Giampetruzzi et al., 2012), has subsequently been found in all the major grapevine-growing areas of the world. It was isolated from plants showing a range of leaf symptoms of varying severity and stunting, even apparently healthy plants, and it also causes significant yield loss. It has been reported that the severity of symptoms depends on the GPGV isolate, the cultivar and environmental parameters, such as the season (Tarquini et al., 2019, 2021). GINV was first reported in Japan (Fan et al., 2017; Giampetruzzi et al., 2012; Yoshikawa et al., 1997), and its biological features and epidemiology are closely similar to those of GPGV. However, GINV seems to be confined to East and North-east Asia (Hily et al., 2020), and is widespread in China (Fan et al., 2017).

An eriophyid mite, *Colomerus vitis*, may be a vector of GPGV (Malagnini et al., 2016). However, the spreading pattern in vineyards and its ability to infect non-*Vitis* weed species, including both herbaceous species such as *Asclepias syriaca*, *Chenopodium albus*, *Rosa*

sp., *Rubus* sp. and *Silene lotiflora* and some woody species, such as *Ailanthus* sp., *Crataegus* sp., *Fraxinus* sp. and *Sambucus* sp. suggests that it may also have additional vectors (Demian et al., 2022; Gualandri et al., 2017; Hily et al., 2021; Malagnini et al., 2016). In addition to local natural spread by vectors, GPGV is disseminated in infected propagating material, which is probably how it has been spread worldwide.

GPGV has been reported from many grapevine-growing regions of the world (Abe & Nabeshima, 2021; Al Rwahnih et al., 2016; Beuve et al., 2015; Czotter et al., 2018; Debat et al., 2020; Eichmeier et al., 2020; Giampetruzzi et al., 2012; Glasa et al., 2014; Jo et al., 2015; Massart et al., 2020; Navrotskaya et al., 2021; Rasool et al., 2019; Reynard et al., 2016; Ruiz-García & Olmos, 2017; Zamorano et al., 2019) and Bertazzon et al. (2016) suggested that the virus might have entered eastern Europe before 2005 and spread subsequently to other European countries. In a recent survey of various grapevine viruses, GPGV was found in German vineyards, but only a single GPGV full-length genomic sequence from German grapevines has been reported to the NCBI GenBank database (Messmer et al., 2021; Reynard et al., 2016).

Knowledge of the time and place of emergence of a crop virus and the evolution of its population is important for understanding its biology, epidemiology and planning its management (Dolan et al., 2018). Hily et al. (2020) reported an analysis of 126 complete sequences of GPGV and concluded that it probably originated in East Asia, especially China. In this study, we confirm and extend these results using phylogenetic, population genetic and dating analyses, especially focusing on the German population of GPGV.

2 | MATERIALS AND METHODS

2.1 | New German sequences

Fifty-six wood samples were collected randomly from different German grapevine-growing regions. They were ground to a fine powder in liquid nitrogen with a cryomill (Retsch). Total RNA was extracted using the Spectrum Total RNA extraction kit (Sigma) as instructed by the manufacturer. RNA quality and quantity were checked by agarose gel electrophoresis and photometric analysis (NanoDrop; Thermo Fisher). The TaqMan RT kit (Applied Biosystems) was used to synthesize first-strand complementary cDNA using random hexamers and oligo(dT). The resulting cDNA was analysed by PCR (Illustra PuReTaq Ready-To-Go PCR Beads; Sigma) using both previously published and newly designed GPGV-specific primer pairs (Table S1) to amplify both the MP-CP coding region and full-length genome sequences. The amplified PCR products were separated using agarose gel electrophoresis and purified using the GenJET gel extraction kit (Thermo Scientific). After ligation into the pJET2.1 cloning vector (Thermo Scientific), the plasmids were used to transform *Escherichia coli* DH5-α competent cells (New England Biolabs). The presence of the target GPGV-fragment was tested by colony PCR using appropriate primer pairs. Positive colonies were cultured, and plasmids were purified using the GenJET plasmid miniprep kit (Thermo Scientific) and sequenced (MWG Eurofins, Germany). The resulting sequences were checked by BLAST searches of the NCBI database. Finally, the sequences were aligned and assembled using Clone Manager 9 software.

2.2 | Sequences and recombination analysis

Gene sequences were manipulated using BioEdit v. 7.2.5. (Hall, 1999). A dataset of the complete coding region of 173 GPGV sequences of various origins retrieved from the NCBI database, together with seven German isolates obtained in this study (OQ533276–OQ533282) and one German isolate from a previous publication (KX522755) (Table S2) were analysed. The sequences were prepared as concatenated ORFs (concats). Then they were aligned using the TranslatorX server (https://translatorx.org/; Abascal et al., 2010) with the MAFFT option (Katoh et al., 2018).

A single sequence (BK011162) of GINV was used as an outlier. One of the two Japanese sequences of GPGV (LC601811) had a region of 85 "unknown" nucleotides out of 7070 nucleotides (1.2%). We used the homologous region of LC601822 for that region.

Seven algorithms (RDP, Chimaera, GENECOV, BootScan, MaxChi, SiScan and 3Seq) in Recombination Detection Program RDP v. 5.5 (Gomez et al., 2018; Martin et al., 2020) were used to screen for possible phylogenetic anomalies resulting from recombination between sequences. A recombination event was considered real if detected by at least four of these algorithms ($p < 10^{-6}$; Ben Mansour et al., 2023). After removing these recombinant isolates, a dataset with 175 nonrecombinant sequences was obtained for further analysis.

An alignment was also made of the MP/CP coding regions of the available sequences ($n = 243$), including 15 sequences isolated from non-*Vitis* hosts and an extra 48 German partial sequences (Table S2). These were aligned as described above and trimmed to a final length of 1517 nucleotides using BioEdit v. 7.2.5.

2.3 | Phylogenetic analysis and dating

Maximum-likelihood (ML) phylogenies of the alignments were constructed using the MEGA X or PhyML programs (Guindon & Gascuel, 2003; Kumar et al., 2018), and the translation model (T93 + γ + I) predicted to be the most appropriate by the MEGA X program (Kumar et al., 2018). Statistical support for phylogenies was tested using 1000 bootstrap replicates or by the SH method (Shimodaira & Hasegawa, 1999).

TempEst v. 1.5.3 was used to assess the temporal signal and phylogenetic coherence of the sequences (Rambaut et al., 2016). A time-scaled ML phylogenetic tree was constructed using the least squared distance (LSD2) method (To et al., 2016) implemented in IQ-TREE v. 2.0.3. The time-scaled tree was viewed using FigTree

v. 1.3 (http://tree.bio.ed.ac.uk/software/figtree/). The node dates of European GPGV isolates ($n=168$) were collected from the phylogeny nexus file and were plotted as a Node Date graph against their numerical rank to infer graphically when there were changes in the rate of addition of divergences in the phylogeny of the virus in Europe (explained below).

ML phylogenies of the European isolates were calculated for the complete concatenated ORFs and the MP-CP coding region of the same sequences using the MEGA X program (Kumar et al., 2018). The resulting patristic distance matrices were compared using PATRISTIC software (Fourment & Gibbs, 2006).

2.4 | Genetic diversity, differentiation and gene flow

DnaSP v. 6.12 was used to determine the genetic diversity (π) and haplotype diversity (H_d; Librado & Rozas, 2009). Tajima's *D*, Fu and Li's *D* and Fu and Li's *F* tests were used to check the neutrality, or otherwise, of selection (Tajima, 1989).

The gene flow and differentiation were estimated by evaluating the fixation index value (F_{ST}); low genetic differentiation $F_{ST}<0.05$, moderate genetic differentiation $0.05<F_{ST}<0.15$, high genetic differentiation $0.15<F_{ST}<0.25$ and complete genetic differentiation $F_{ST}>0.25$. Gene flow was assessed by the migrant value (N_m), in which $N_m \geq 1$ was considered high gene flow, $0.25<N_m<0.99$ medium gene flow and $N_m<0.249$ low gene flow (Balloux & Lugon-Moulin, 2002; Hudson, 2000; Hudson et al., 1992; Wright, 1978). Three parameters, K_s^*, Z^* and S_{nn}, were used to further challenge the results obtained (Gao et al., 2017; Tajima, 1989; Valouzi et al., 2022).

2.5 | Variability of German GPGV isolates

The investigation of the selection pressure affecting codon sites of the MP and CP genes for the 56 German isolates was assessed using three methods (SLAC, FEL and FUBAR) implemented within the Datamonkey online application, and the codon sites detected by at least two or more of these methods were considered real (Delport et al., 2010).

3 | RESULTS

3.1 | Recombination and phylogeny

The RDP5 analysis of the full-length concatenated ORFs (concats) revealed four recombinants (Table S3); two of Australian origin (OQ199011–OQ199020), one of Italian origin (MH087443) and one of German origin (OQ533279), which was detected with a low probability RDP ($p=1.943\times10^{-4}$), GENECOV (4.877×10^{-4}), MaxChi (1.931×10^{-5}), Chimaera (5.919×10^{-5}), SiScan (1.227×10^{-5}) and 3Seq (2.500×10^{-2}). The summary ML phylogenetic tree (Figure 1) of

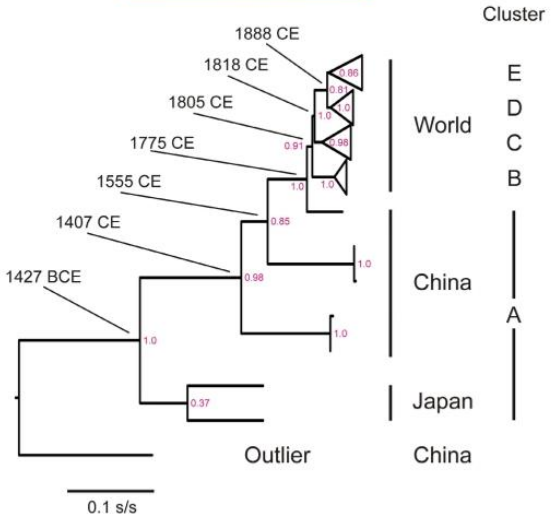

FIGURE 1 A summary of a maximum-likelihood phylogenetic tree of 175 grapevine Pinot gris virus (GPGV) concats (concatenated ORFs) and one grapevine berry inner necrosis virus (GINV) (BK011162) concat as an outlier. The branch to the outlier is drawn 0.1% of its calculated length, and those to the cluster of two Japanese isolates (LC601811/2) at the base of the GPGV phylogeny are drawn 1% of their estimated length. The scale bar shows the evolutionary distance as substitutions per site (s/s). The SH estimates of statistical significance of each node are shown in red. Cluster A includes seven Asian GPGV isolates. Four collapsed clusters (B–E) comprise B containing two isolates, C containing 16 isolates, D containing 71 isolates and E containing 79 isolates. [Colour figure can be viewed at wileyonlinelibrary.com]

the 176 nonrecombinant concats, including an outlier, groups them into five statistically supported Clusters A–E; a basal paraphyletic group (A) of concats only from Asian isolates and four monophyletic groups (B–E) of concats from European and world isolates. Figure 2 shows details of the 168 European and world concats together with a Chinese concat as outlier. Cluster C included concats from isolates from Germany (4/7), Italy (8/31), France (1/51), China (1/8) and Australia (2/25). Cluster B contained only two Italian isolates. Cluster D contained all the Russian and Slovakian isolates, most of the Australian (23/25) and the Italian (17/31), but few German (2/7), American (1/11) and Canadian (3/13) isolates. In contrast, almost all French (52/53), American (10/11), Canadian (10/13) and few German (1/7), Italian (4/31) and Chinese (2/8) isolates grouped within Cluster E. Note that the relative positions of Clusters A, B and C are slightly different in Figures 1 and 2, as a consequence of using different concats as an outlier.

Some GPGV isolates are represented in GenBank only by their MP and CP regions, so we checked by the PATRISTIC method (Fourment & Gibbs, 2006) how representative the phylogeny of the MC/CPs was of the phylogeny of the complete concats using the 168 sequences of Clusters B to E with the nearest Chinese sequence as an outlier. Figure 3 shows a graph comparing pairwise the patristic distances in the two phylogenies. Although they are significantly

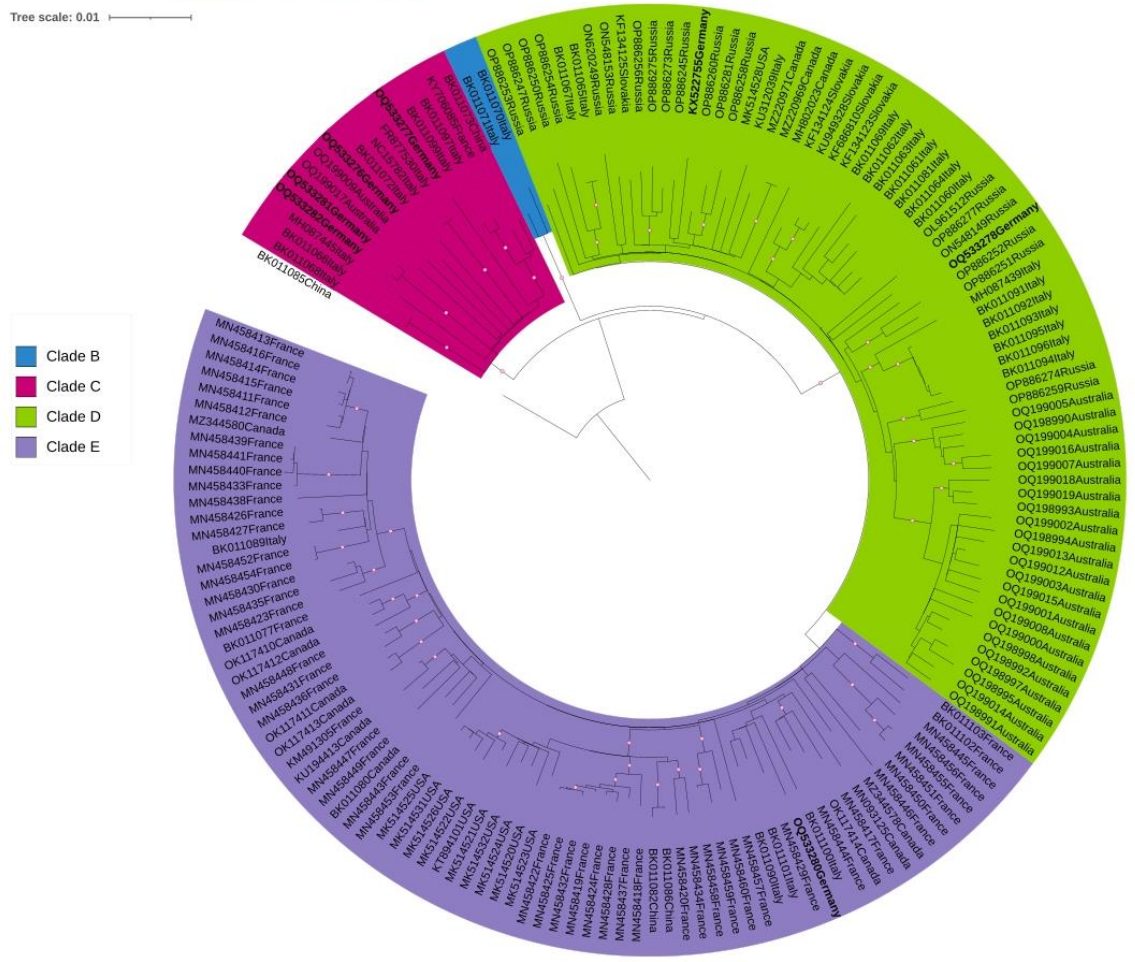

FIGURE 2 Circular maximum-likelihood phylogenetic tree of 169 complete sequences of grapevine Pinot gris virus (GPGV) isolates, including seven German isolates (highlighted in bold). Colours indicate different clusters. Bootstrap values close to 100% are indicated at the tree nodes. Scale is 0.01 substitutions/site.Tree was viewed using iTOL. [Colour figure can be viewed at wileyonlinelibrary.com]

correlated, the relationship is broad ($r = 0.657$). Inspection of the phylogenies shows that whereas Clusters D and E are distinct in the complete concat phylogeny, they form one large cluster in the MP-CP phylogeny (Figure 4). Nonetheless, Clusters B and C are still statistically distinct in the MP-CP phylogeny, and two German isolates have joined the two Italian isolates in Cluster C. Figure 4, where the German isolates are highlighted, shows that they are present in almost all clusters except Cluster A.

3.2 | Dating

A TempEst analysis of the complete GPGV concats for which accession dates are available was done. These comprised 149 sequences, including the two Japanese basal isolates but not the five basal Chinese isolates. Nonetheless, these sequences showed a significant temporal signal indicating a most recent common ancestor (MRCA) dated around 1285 CE ($p = 0.01–0.001$). A time-scaled tree was then calculated using IQ-TREE v. 2.2.2, which includes the Least Squared Dating method of To et al. (2016), which, unlike TempEst, allows sequences of unknown accession dates to be included, and extends dating to them. The time-scaled tree constructed using 175 nonrecombinant GPGV concats and a GINV isolate (BK011062) used as an outlier is summarized in Figure 5. The basal node of the phylogeny has concats from a pair of Japanese isolates obtained from the ornamental vine, *Vitis coignetiae*, as the sister lineage to the concats of all the other isolates of GPGV, of which the five most basal are from China. A single branch connects the basal Cluster A to Clusters B to E, which are of the concats of 168 isolates from other parts of the world, and these are shown in this figure as collapsed

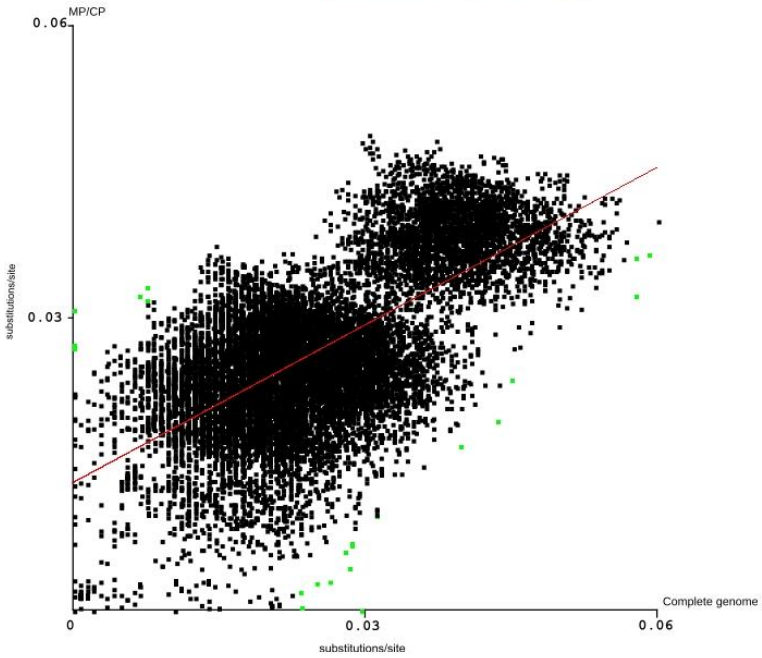

FIGURE 3 Comparison of pairwise patristic distances in maximum-likelihood phylogenies of movement protein (MP)/coat protein (CP) and complete genome of grapevine Pinot gris virus (GPGV). Green points indicate paired distances over two standard deviations from the mean difference and the red linear regression line has a correlation coefficient $r=0.677$. [Colour figure can be viewed at wileyonlinelibrary.com]

triangles. The pair of Japanese isolates from *V. coignetiae* diverged from the others in 1427 BCE, 3500 years ago, and much earlier than the TempEst analysis indicated. The earliest divergence of the five Chinese isolates obtained from *V. vinifera* was in 1407 CE (only 600 years ago). The "world isolates" in Clusters B to E diverged from the nearest Chinese isolate in 1775 CE, 1805 CE, 1818 CE and 1808 CE, respectively. Thus, GPGV may have travelled to Europe between 1775 and 1805 CE, before spreading further afield in the succeeding 220 years.

Figure 6 shows a graph of the numerical dates of all the nodes and twigs in the IQ-TREE ML phylogeny of Clusters B to E. The earliest node was dated 1805 CE, and the most recent was 2021 CE, a total of 218 years. If nodes in the phylogeny had appeared randomly, then one would expect the curve of ranked dates to follow a simple quadratic. However, there are regions of the curve that are flattened, and these are periods where new nodes are appearing more frequently than would be expected by chance. These regions are dated around 1870–1880 CE (A), 1925–1930 CE (B) and 1950 CE (C).

3.3 | Genetic diversity, gene flow and genetic differentiation analysis of GPGV isolates

The DNaSP. v. 6.12 program was used to estimate the genetic diversity of the GPGV isolates using their MP and CP coding regions (Table S4). These analyses showed that members of Cluster A had the greatest genetic diversity for both MP ($\pi=0.11592\pm0.004$) and CP ($\pi=0.104\pm0.0033$). Similarly, isolates from Asia had the highest genetic diversity (MP, $\pi=0.097\pm0.002$ and CP, $\pi=0.089\pm0.002$), followed by the European population with genetic diversity (MP, $\pi=0.020\pm0.000001$ and CP, $\pi=0.021\pm0.000004$). However, the population from the American continent had the lowest genetic diversity (MP, $\pi=0.012\pm0.00007$ and CP, $\pi=0.016\pm0.0005$).

A comparison between populations from different European countries indicated that the Slovakian population had the lowest genetic diversity for both genes (MP, $\pi=0.011\pm0.0006$ and CP, $\pi=0.010\pm0.0006$). In contrast, the German population, the subject of our study, had the greatest genetic diversity for both genes (MP, $\pi=0.023\pm0.00001$ and CP, $\pi=0.025\pm0.00002$) compared to all other European GPGV populations. The number of haplotypes (H) of the 56 tested German isolates based on the MP gene was $H=47$, while for the CP gene was $H=44$. The haplotype diversity was high for both genes ($H_d>0.98$) for all the analysed worldwide populations.

The statistical tests Tajima's *D*, Fu and Li's *D* and Fu and Li's *F* on the MP and the CP genes distinguish which populations have been evolving neutrally from those that have been evolving under non-random processes such as population selection and expansion or contraction. All continental populations were negative. The statistical tests were significant for the European population (both genes) and the American population (only for the CP gene). But the populations from the Australian and Asian continents, and for the German population, were not significantly non-random.

We also determined the total genetic variation within a population relative to the total population based on major phylogroups and geographical origins. It was estimated from *F* statistics (F_{ST}) that the populations from different phylogroups were genetically distinct from each other (Table S5), with a high fixation index $F_{ST}>0.23$ measured for both genes. However, the N_m value represents the gene flow between populations ranging between high values of $N_m>1$, indicating a frequent gene flow observed between members of

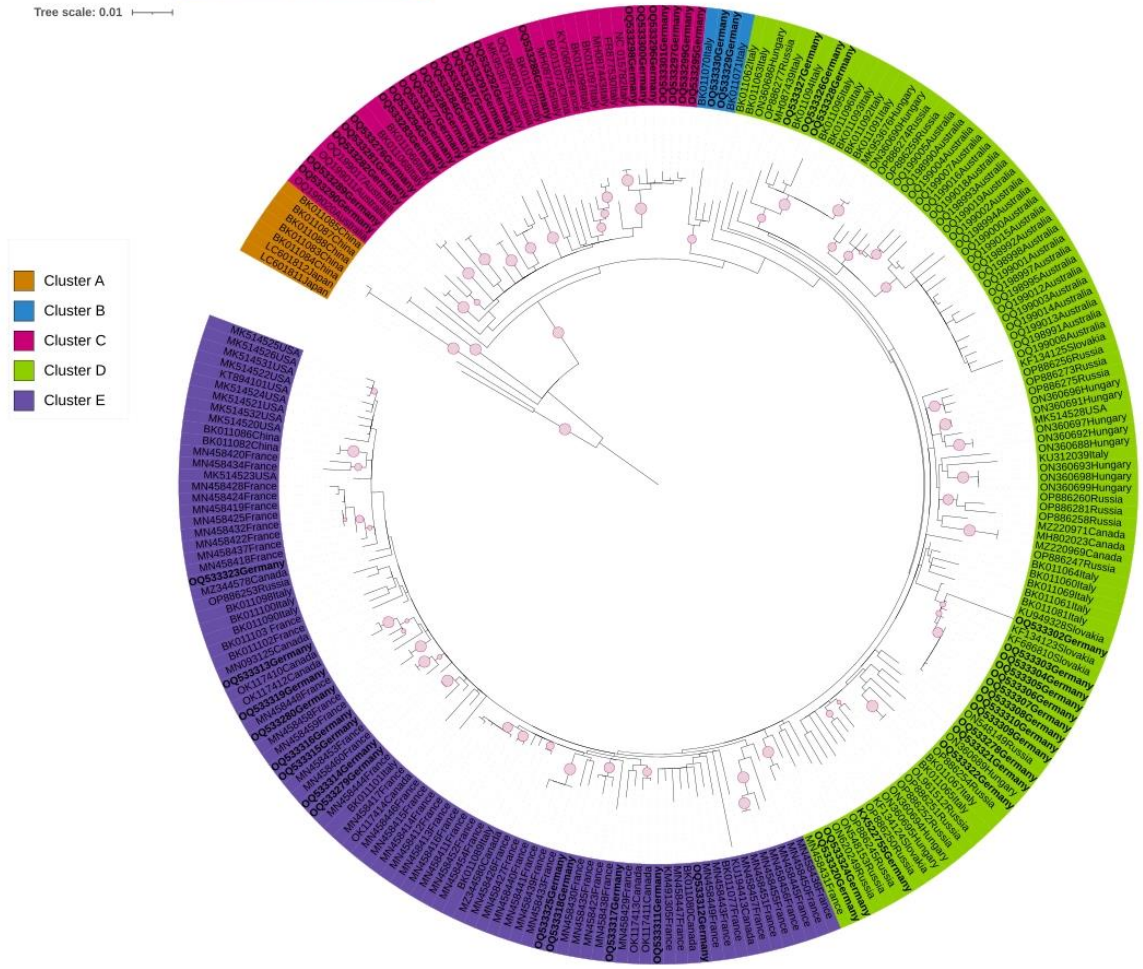

FIGURE 4 Circular maximum-likelihood phylogenetic tree of 243 grapevine Pinot gris virus (GPGV) isolates, including German isolates (highlighted in bold) calculated from their movement protein (MP)/coat protein (CP) sequences. Colours indicate different clusters. Bootstrap values close to 100% are indicated at the tree nodes. Scale is 0.01 substitutions/site. Tree was viewed using iTOL. [Colour figure can be viewed at wileyonlinelibrary.com]

Cluster A (Asian isolates) and the remaining clusters. A low value of $N_m < 1$, indicating infrequent gene flow, was observed between members from the remaining clusters.

F statistics were also used to check whether the German and other populations from different geographical origins were genetically distinct (Table S5). The German isolates were genetically similar to the Italian population ($F_{ST} < 0.03$) for both genes. However, it was genetically distinct from all remaining worldwide populations (France, Slovakia, Russia, Hungary, China, Japan, United States, Canada and Australia), with an $F_{ST} > 0.10$. The N_m value showed a frequent gene flow of GPGV between the German population and different countries ($N_m > 1$). In contrast, there was infrequent gene flow only between German and Japanese populations ($N_m < 1$). It is worth noting that the *p* value for all permutation-based tests was statistically significant between the German population and almost all worldwide populations, except for the Slovakian population, which was nonsignificant.

3.4 | Variability of German GPGV isolates

The individual codons that were under selection were identified using three methods (SLAC, FEL and FUBAR) implemented within the Datamonkey online application (Delport et al., 2010); the codon sites detected by at least two of these methods were considered real (Table S6). The resulting nonsynonymous-to-synonymous nucleotide diversity ratios of the different GPGV populations showed that all are under strong negative selection for both genes (dN/dS < 0.15); 48 and 25 codon sites under negative selection for MP and CP, respectively. Only six codon sites were under positive selection

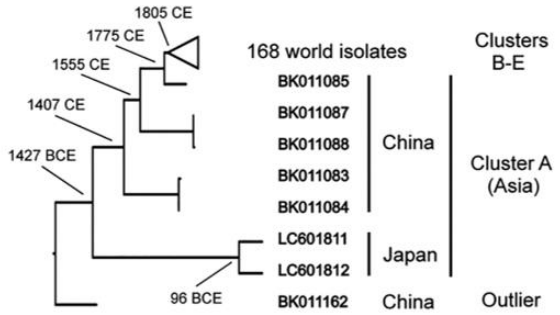

FIGURE 5 A summary of the basal regions of the phylogeny of the concats of 175 grapevine Pinot gris virus (GPGV) sequences and one grapevine berry inner necrosis virus (GINV) (BK011162) concat as an outlier. The branch to the outlier is drawn 2.0% of its calculated length, and those to the cluster of two Japanese isolates (LC601811/2) at the base of the GPGV phylogeny are drawn 20.0% of their calculated length. All the nodes shown had SH statistical support greater than 0.80 (range 1–0.81, mean 0.96). The collapsed clusters B–E of 168 "world isolates" of GPGV mostly came from Europe, North America and Australia, and include three more from China. The estimated dates of some nodes are shown, and all are analysed (see below). [Colour figure can be viewed at wileyonlinelibrary.com]

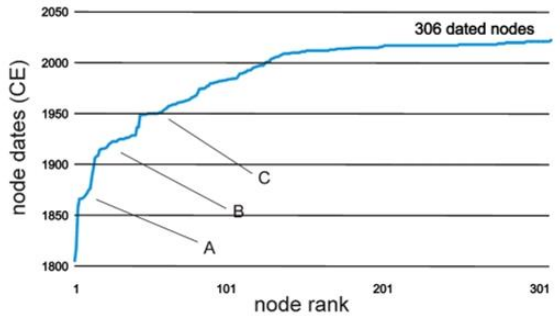

FIGURE 6 Graph of the dates of the nodes and twigs of the 168 concat sequences of isolates forming Clusters B–E, dating from 1805 CE to 2021 CE. Regions of the curve marked A, B and C are where new nodes appeared more frequently than would be expected from the general trend of the curve (i.e., the curve flattened; see text). [Colour figure can be viewed at wileyonlinelibrary.com]

pressure; five are located at the 3′ end of the MP (Figure S1), and one codon site in the CP gene (Figure S2).

4 | DISCUSSION

Our study aimed to update the phylogeny of the world population of GPGV, both in space and time and to place the German population of the virus in that world context. In 2020 similar studies of GPGV populations by Hily et al. (2020) and Tokhmechi et al. (2021) concluded that GPGV emerged in China in the mid-20th century and spread to Europe, initially Italy, and from there, it spread to the rest of the world, and even back to China. Recently, more GPGV sequences have been stored in GenBank, and particularly significant for phylogenetic dating are those from two isolates of GPGV from Japan found in *V. coignetiae*, a wild grapevine of North-east Asia (Abe & Nabeshima, 2021), that we found to be basal in the GPGV phylogeny. Therefore, this result enabled us to determine with more certainty the origin of GPGV and the dating of the basal regions of its phylogeny, especially using the LSD application in IQ-TREE that allows the inclusion of the undated early sequences from China.

The genus *Vitis* consists of 60–80 ecospecies of vine (Anon, 2023; Emanuelli et al., 2013; Hardie, 2000), probably all interfertile, and found throughout the northern temperate zone of both Old and New Worlds. The most widely cultivated grapevine species is *V. vinifera* subsp. *vinifera*, which was domesticated in Eurasia over the past 11,000–7000 years (Allaby, 2023; Dong et al., 2023; Myles et al., 2011), but the centre of diversity of the genus is East Asia and North America. *V. coignetiae* is found wild in Korea, Japan and the Russian Far East but is now grown worldwide as the ornamental Crimson Glory vine (Myles et al., 2011; Nakagawa et al., 1991). It produces small fruits with large seeds; fermented drinks flavoured by these and other fruits have been made in East Asia for at least 4000 years (Anon, 2021). Wild grapes were also used for medicinal purposes in Japan more than 10,000 years ago (Anon, 2020), and hybrids of *V. vinifera* and *V. coignetiae* (called *yamabudou* in Japanese) and other species were successfully produced (Yamashita & Mochioka, 2014). It was not until the Han Dynasty (226 BCE–221 CE) that *V. vinifera* subsp. *vinifera* grapes for winemaking were first introduced to China, and their initial use peaked there in the 13th and 14th centuries (Li et al., 2018). Thus, the basal node of the GPGV phylogeny (1427 BCE) linking GPGV in China and Japan occurred before *V. vinifera* subsp. *vinifera* vines had been taken to China, whereas the post-1407 CE divergence of GPGV in China occurred after the more productive Eurasian grapevines had been introduced, and so the divergences recorded by the later branchings probably occurred in European grapevines. It is noteworthy that GINV, the closest known relative of GPGV, has only been isolated from plants in North-east Asia, which supports our conclusion that the GPGV-GINV lineage emerged in that region.

Most isolates of GPGV obtained from *V. vinifera* subsp. *vinifera* worldwide form four clusters (B–E) dated after 1805 CE. We found that a graph of the node dates for these clusters gave an irregular curve with obvious flattenings around 1870 CE–1880 CE, 1920 CE–1930 CE and around 1950 CE, marked A–C in Figure 6. Flattening of the curve occurs when several nodes with the same or similar dates appear, whereas a smooth curve would be expected if the appearance of nodes was random and therefore followed a simple quadratic. Therefore flattenings may be associated with periods when the transport of infected grapevine materials may have increased. The first flattening of the curve (A) corresponds to the date of the devastating European phylloxera epidemic caused by the importation from North America of *Daktulosphaira vitifoliae*, a hemipterous insect that feeds on the root of vines. Phylloxera was controlled by grafting *Vitis* scions onto selected phylloxera-resistant *Vitis*

rootstocks, mostly obtained from North America, rather than growing them on their own roots, which had been the practice. This procedure must have greatly increased the human intervention in grape crops, and spread by mechanical means and vectors, and hence an increased appearance of new nodes in the GPGV phylogeny.

In 1920–1930, marked by the second flattering of the node date curve (B), there were worldwide events that may have affected grape production, such as the prohibition of alcohol in the United States, which may have greatly impacted grape production (Alston & Sambucci, 2019) and caused a change to table grape production and therefore the transport of grapevine rootstocks and scions. This period also coincided with the Great Depression in Europe, which affected consumer and agriculture preferences, causing a decline in the demand for wine, considered a luxury product and therefore affecting grape production (Federico, 2005). The third flattening of the node date curve observed in the 1950s occurred when many countries adopted regulatory laws and hybrid varieties (Hajdu, 2015; Reynolds, 2015; Ruehl et al., 2015). The second and third flattening of the node date curve (B and C), in 1920–1930 and 1950, may also merely reflect efforts after World War I and II to expand the remnants of the crop in Europe and to expand viticulture in other parts of the world, such as the Americas and Australia.

It is also noteworthy that the nodes dated 1870–1880 CE were to branches that consisted of 80% European isolates, the 1925–1930 CE nodes to 40% European isolates and the 1950 CE nodes to 30% European isolates, whereas they showed the reverse relationship to Australian isolates (10%, 29% and 54%, respectively), which is a clear indication that GPGV spread first to, and within, Europe and from there to other parts of the world. The *F* statistics showed that European GPGV populations were genetically distinct from other continental populations, but perhaps also the oldest. However, a continuous gene flow ($N_m > 1$) was observed between them, suggesting a dynamic evolutionary scenario aiming to reduce their genetic divergence and maintain a certain level of gene connectivity between them. This result also agrees with the observed neutrality tests, which were negative and significant for the European population, suggesting a recent population growth.

We have shown that German GPGV isolates were found in four clusters (B, C, D and E). The German isolates were genetically more diverse, for both genes, than those of other European populations, suggesting that the initial stages of GPGV's invasion of Europe were in Germany, not Italy (Hily et al., 2020; Tokhmechi et al., 2021). In contrast, most GPGV isolates from other regions (Australia, United States, France, Russia, Japan and China) tended to cluster according to their origin, as previously described (Elçi et al., 2018; Fajardo et al., 2017; Hily et al., 2020; Tokhmechi et al., 2021). The German population was found to have a low degree of genetic differentiation and a high level of gene flow with the Italian population, which agrees with the observed dispersion of German and Italian isolates within the phylogenetic tree. They were found together in each cluster. Such gene flow between German and Italian populations could be explained by a long history of exchanging plant material for grafting and international breeding programmes between these two countries (Bavaresco et al., 2015).

All the neutrality tests for the German population were negative, indicating a polymorphism. Still, they were nonsignificant for both genes, suggesting that the polymorphism frequency was less than expected. The evolutionary selection pressure ratio was estimated for both genes in the German GPGV population, indicating a negative selection, which is a purifying process that removes isolates with deleterious mutations and decreases the frequency of less-fit viral variants (García-Arenal et al., 2003; Hughes, 2009). The investigation of the site-specific selection pressure affecting the MP and CP genes of the German population revealed the presence of 48 and 25 codon sites detected to be under negative selection for MP and CP, respectively. Only six codon sites were assigned for a positive selection pressure (MP: five codon sites located at the 3′ end, CP: one codon site). Notably, sites under positive selection were present more often in the MP than in the CP. Positive selection of the MP gene might support the spread of beneficial virus variants by promoting systemic infection (Yoshikawa et al., 2006; Zhang, 2008).

Our study updated the likely origin of GPGV after a previous report suggested that it originated in China. Newly available sequences in the GenBank led us to conclude that Japan, and presumably other regions of North-east Asia where *V. coignetiae* originated, is the centre of emergence of this virus, and that it may have moved later to domesticated grape (*V. vinifera*). The fact that infected vines may appear healthy, together with lax regulations and lack of testing for GPGV, favoured the easy spread of the virus internationally and multiple reintroductions locally, as can be concluded from the observed gene flow between continental populations.

ACKNOWLEDGEMENTS

This study was funded by the Ministerium für Wirtschaft, Verkehr, Landwirtschaft und Weinbau, Rhineland-Palatinate, Germany. The study was supported by the Ministry of Agriculture of the Czech Republic, institutional support MZE-RO0423.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon request.

ORCID

Karima Ben Mansour https://orcid.org/0000-0002-0054-5954

REFERENCES

Abascal, F., Zardoya, R. & Telford, M.J. (2010) TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research*, 38, W7–W13.

Abe, J. & Nabeshima, T. (2021) First report of grapevine Pinot gris virus in wild grapevines (*Vitis coignetiae*) in Japan. *Journal of Plant Pathology*, 103, 725.

Al Rwahnih, M., Alabi, O.J., Hwang, M.S., Tian, T., Mollov, D. & Golino, D. (2021) Characterization of a new nepovirus infecting grapevine. *Plant Disease*, 105, 1432–1439.

Al Rwahnih, M., Golino, D. & Rowhani, A. (2016) First report of grapevine Pinot gris virus infecting grapevine in the United States. *Plant Disease*, 100, 1030.

Allaby, R.G. (2023) Two domestications for grapes. *Science*, 379, 880–881.

Alston, J. & Sambucci, O. (2019) Grapes in the world economy. In: Cantu, D. & Walker, M.A. (Eds.) *The grape genome. Compendium of plant genomes*. Cham: Springer, pp. 1–24.

Anon. (2020) Inaizumi nōen no yama budō ["Wild grapes" from Inaizumi farm]. Available from: https://inaizumi-farm.com/news/5eb58c3172b9117f0edc781d [Accessed 22nd July 2023]

Anon. (2021) *Vitis coignetiae*. Available from: https://en.wikipedia.org/wiki/Vitis_coignetiae [Accessed 31st July 2023].

Anon. (2023) *Vitis*. Available from: https://en.wikipedia.org/wiki/Vitis [Accessed 21st July 2023].

Balloux, F. & Lugon-Moulin, N. (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, 11, 155–165.

Bavaresco, L., Gardiman, M., Brancadoro, L., Espen, L., Failla, O., Scienza, A. et al. (2015) Grapevine breeding programs in Italy. In: Reynolds, A.G. (Ed.) *Grapevine breeding programs for the wine industry*. Amsterdam: Woodhead Publishing, pp. 135–157.

Ben Mansour, K., Gibbs, A.J., Komínková, M., Komínek, P., Brožová, J., Kazda, J. et al. (2023) Watermelon mosaic virus in the Czech Republic, its recent and historical origins. *Plant Pathology*, 72, 1528–1538.

Bertazzon, N., Filippin, L., Forte, V. & Angelini, E. (2016) Grapevine Pinot gris virus seems to have recently been introduced to vineyards in Veneto, Italy. *Archives of Virology*, 161, 711–714.

Beuve, M., Candresse, T., Tannières, T. & Lemaire, O. (2015) First report of grapevine Pinot gris virus (GPGV) in grapevine in France. *Plant Disease*, 99, 293.

Czotter, N., Molnar, J., Szabó, E., Demian, E., Kontra, L., Baksa, I. et al. (2018) NGS of virus-derived small RNAs as a diagnostic method used to determine viromes of Hungarian vineyards. *Frontiers in Microbiology*, 9, 122.

Debat, H., Facundo, L., Moyano, S., Zavallo, D., Asurmendi, S. & Gomez-Talquenca, S. (2020) First report of grapevine Pinot gris virus infecting grapevine in Argentina. *Journal of Plant Pathology*, 102, 1321.

Delport, W., Poon, A.F.Y., Frost, S.D.W. & Kosakovsky Pond, S.L. (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*, 26, 2455–2457.

Demian, E., Jaksa-Czotter, N. & Varallyay, E. (2022) Grapevine Pinot gris virus is present in different non-*Vitis* hosts. *Plants*, 11, 1830.

Dolan, P.T., Whitfield, Z.J. & Andino, R. (2018) Mechanisms and concepts in RNA virus population dynamics and evolution. *Annual Review of Virology*, 5, 69–92.

Dong, Y., Duan, S., Xia, Q., Liang, Z., Dong, X., Margaryan, K. et al. (2023) Dual domestications and origin of traits in grapevine evolution. *Science*, 379, 892–901.

Eichmeier, A., Penazova, E. & Nebish, A. (2020) First report of grapevine Pinot gris virus on grapevines in Armenia. *Plant Disease*, 104, 1000.

Elçi, E., Gazel, M., Roumi, V. & Çağlayan, K. (2018) Incidence, distribution and limited genetic variability among Turkish isolates of grapevine Pinot gris virus from different grapevine cultivars. *Journal of Plant Diseases and Protection*, 125, 469–476.

Emanuelli, F., Lorenzi, S., Grzeskowiak, L., Catalano, V., Stefanini, M., Troggio, M. et al. (2013) Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. *BMC Plant Biology*, 13, 39.

Fajardo, T.V.M., Eiras, M. & Nickel, O. (2017) First report of grapevine Pinot gris virus infecting grapevine in Brazil. *Australasian Plant Disease Notes*, 12, 2015–2018.

Fan, X.D., Zhang, Z.P., Ren, F., Hu, G.J., Zhou, J., Li, Z.N. et al. (2017) Occurrence and genetic diversity of grapevine berry inner necrosis virus from grapevines in China. *Plant Disease*, 101, 144–149.

Federico, G. (2005) Not guilty? Agriculture in the 1920s and the great depression. *Journal of Economic History*, 65, 949–976.

Fourment, M. & Gibbs, M.J. (2006) PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. *BMC Evolutionary Biology*, 6, 1.

Fuchs, M. (2020) Grapevine viruses: a multitude of diverse species with simple but overall poorly adopted management solutions in the vineyard. *Journal of Plant Pathology*, 102, 643–653.

Gao, F., Zou, W., Xie, L. & Zhan, J. (2017) Adaptive evolution and demographic history contribute to the divergent population genetic structure of potato virus Y between China and Japan. *Evolutionary Applications*, 10, 379–390.

García-Arenal, F., Fraile, A. & Malpica, J.M. (2003) Variation and evolution of plant virus populations. *International Microbiology*, 6, 225–232.

Giampetruzzi, A., Roumi, V., Roberto, R., Malossini, U., Yoshikawa, N., La Notte, P. et al. (2012) A new grapevine virus discovered by deep sequencing of virus- and viroid-derived small RNAs in cv Pinot Gris. *Virus Research*, 163, 262–268.

Glasa, M., Predajňa, L., Komínek, P., Nagyová, A., Candresse, T. & Olmos, A. (2014) Molecular characterization of divergent grapevine pinot gris virus isolates and their detection in Slovak and Czech grapevines. *Archives of Virology*, 159, 2103–2107.

Gomez, C., Gonzalez, G. & Meneses, E. (2018) Algorithm for history reconstruction of viral recombination events: preliminary results. *2018 IEEE international work conference on bioinspired intelligence, IWOBI 2018 - proceedings*. https://doi.org/10.1109/IWOBI.2018.8464134

Gualandri, V., Asquini, E., Bianchedi, P., Covelli, L., Brilli, M., Malossini, U. et al. (2017) Identification of herbaceous hosts of the grapevine Pinot gris virus (GPGV). *European Journal of Plant Pathology*, 147, 21–25.

Guindon, S. & Gascuel, O.A. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52, 696–704.

Hajdu, E. (2015) Grapevine breeding in Hungary. In: Reynolds, A.G. (Ed.) *Grapevine breeding programs for the wine industry*. Amsterdam: Woodhead Publishing, pp. 103–134.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95–98.

Hardie, W.J. (2000) Grapevine biology and adaptation to viticulture. *Australian Journal of Grape and Wine Research*, 6, 74–81.

Hily, J.M., Komar, V., Poulicard, N., Vigne, E., Jacquet, O., Protet, N. et al. (2021) Biological evidence and molecular modeling of a grapevine Pinot gris virus outbreak in a vineyard. *Phytobiomes Journal*, 5, 464–472.

Hily, J.M., Poulicard, N., Candresse, T., Vigne, E., Beuve, M., Renault, L. et al. (2020) Datamining, genetic diversity analyses, and phylogeographic reconstructions redefine the worldwide evolutionary history of grapevine Pinot gris virus and grapevine berry inner necrosis virus. *Phytobiomes Journal*, 4, 165–177.

Hudson, R.R. (2000) A new statistic for detecting genetic differentiation. *Genetics*, 155, 2011–2014.

Hudson, R.R., Boos, D.D. & Kaplan, N.L. (1992) A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution*, 9, 138–151.

Hughes, A. (2009) Small effective population sizes and rare nonsynonymous variants in potyviruses. *Virology*, 393, 127–134.

Jo, Y., Choi, H., Cho, J.K., Yoon, J.Y., Choi, S.K. & Cho, W.K. (2015) In silico approach to reveal viral populations in grapevine cultivar Tannat using transcriptome data. *Scientific Reports*, 5, 15841.

Katoh, K., Rozewicki, J. & Yamada, K.D. (2018) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20, 1160–1166.

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35, 1547–1549.

Li, H., Wang, H., Li, H., Goodman, S., van der Lee, P., Xu, Z. et al. (2018) The worlds of wine: old, new and ancient. *Wine Economics and Policy*, 7, 178–182.

Librado, P. & Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, 1451–1452.

Malagnini, V., de Lillo, E., Saldarelli, P., Beber, R., Duso, C., Raiola, A. et al. (2016) Transmission of grapevine Pinot gris virus by *Colomerus vitis* (Acari: Eriophyidae) to grapevine. *Archives of Virology*, 161, 2595–2599.

Martin, D.P., Varsani, A., Roumagnac, P., Botha, G., Maslamoney, S., Schwab, T. et al. (2020) RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evolution*, 7, veaa087.

Massart, S., Vankerkoven, L., Blouin, A.G., Nourinejhad Zarghani, S. & Wetzel, T. (2020) First report of grapevine Pinot gris virus and grapevine rupestris stem pitting-associated virus in grapevine in Belgium. *Plant Disease*, 104, 1879.

Messmer, N., Bohnert, P., Schumacher, S. & Fuchs, R. (2021) Studies on the occurrence of viruses in planting material of grapevines in southwestern Germany. *Viruses*, 13, 2481–11.

Myles, S., Boyko, A.R., Owens, C.L., Brown, P.J., Grassi, F., Aradhya, M.K. et al. (2011) Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 3530–3535.

Nakagawa, S., Horiuchi, S., Matsui, H., Yuda, E., Yamada, S., Murai, Y. et al. (1991) Distribution and leaf morphology of wild grapes native to Japan. *Journal of the Japanese Society for Horticultural Science*, 60, 31–39.

Navrotskaya, E., Porotikova, E., Yurchenko, E., Galbacs, Z.N., Varallyay, E. & Vinogradova, S. (2021) High-throughput sequencing of small RNAs for diagnostics of grapevine viruses and viroids in Russia. *Viruses*, 13, 2432.

Rambaut, A., Lam, T.T., Carvalho, L.M. & Pybus, O.G. (2016) Exploring the temporal structure of heterochronous sequences using TempEst (formerly path-O-gen). *Virus Evolution*, 2, vew007.

Rasool, S., Naz, S., Rowhani, A., Diaz-Lara, A., Golino, D.A., Farrar, K.D. et al. (2019) Survey of grapevine pathogens in Pakistan. *Journal of Plant Pathology*, 101, 725–732.

Reynard, J.S., Schumacher, S., Menzel, W., Fuchs, J., Bohnert, P., Glasa, M. et al. (2016) First report of grapevine Pinot gris virus in German vineyards. *Disease Notes*, 100, 2545.

Reynolds, A.G. (2015) Grapevine breeding in France – a historical perspective. In: Reynolds, A.G. (Ed.) *Grapevine breeding programs for the wine industry*. Amsterdam: Woodhead Publishing, pp. 65–76.

Ruehl, E., Schmid, J., Eibach, R. & Töpfer, R. (2015) Grapevine breeding programmes in Germany. In: Reynolds, A.G. (Ed.) *Grapevine breeding programs for the wine industry*. Amsterdam: Woodhead Publishing, pp. 77–101.

Ruiz-García, A.B. & Olmos, A. (2017) First report of grapevine Pinot gris in grapevine in Spain. *Plant Disease*, 101, 1070.

Shimodaira, H. & Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, 16, 1114–1116.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585–595.

Tarquini, G., Ermacora, P. & Firrao, G. (2021) Polymorphisms at the 3′ end of the movement protein (MP) gene of grapevine Pinot gris virus (GPGV) affect virus titre and small interfering RNA accumulation in GLMD disease. *Virus Research*, 302, 198482.

Tarquini, G., Zaina, G., Ermacora, P., de Amicis, F., Franco-Orozco, B., Loi, N. et al. (2019) Agroinoculation of grapevine Pinot gris virus in tobacco and grapevine provides insights on viral pathogenesis. *PLoS One*, 14, e0214010.

To, T.H., Jung, M., Lycett, S. & Gascuel, O. (2016) Fast dating using least-squares criteria and algorithms. *Systematic Biology*, 65, 82–97.

Tokhmechi, K., Abadkhah, M. & Koolivand, D. (2021) Emerging and population analysis of grapevine Pinot gris virus isolates from Iran. *3 Biotech*, 11, 368.

Valouzi, H., Shahmohammadi, N., Golnaraghi, A., Moosavi, M.R. & Ohshima, K. (2022) Genetic diversity and evolutionary analyses of potyviruses infecting narcissus in Iran. *Journal of Plant Pathology*, 104, 237–250.

Wright, S. (1978) *Evolution and the genetic of population*, Vol. 4, Variability within and among natural populations. Chicago: University of Chicago Press.

Yamashita, H. & Mochioka, R. (2014) Wild grape germplasms in Japan. *Advances in Horticultural Science*, 28, 214–224.

Yoshikawa, N., Iida, H., Goto, S., Magome, H., Takahashi, T. & Terai, Y. (1997) Grapevine berry inner necrosis, a new trichovirus: comparative studies with several known trichoviruses. *Archives of Virology*, 142, 1351–1363.

Yoshikawa, N., Saitou, Y., Kitajima, A., Chida, T., Sasaki, N. & Isogai, M. (2006) Interference of long-distance movement of grapevine berry inner necrosis virus in transgenic plants expressing a defective movement protein of apple chlorotic leaf spot virus. *Phytopathology*, 96, 378–385.

Zamorano, A., Medina, G., Fernández, C., Cui, W., Quiroga, N. & Fiore, N. (2019) First report of grapevine Pinot gris virus in grapevine in Chile. *Plant Disease*, 103, 1438.

Zhang, J. (2008) Positive selection, not negative selection, in the pseudogenization of *rcsA* in *Yersinia pestis*. *Proceedings of the National Academy of Sciences of the United States of America*, 105, E69.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ben Mansour, K., Gibbs, A.J., Meßmer, N., Fuchs, R., Wetzel, T. & Winterhagen, P. (2024) Grapevine Pinot gris virus in Germany: From where did the virus come, and when? *Plant Pathology*, 73, 455–464. Available from: https://doi.org/10.1111/ppa.13824

III. Study 3

-Title: Multiple infections with viruses of the family *Tymoviridae* in Czech grapevines

-Brief description: This study focused on the viruses of the *Tymoviridae* family that infect grapevines in the Czech Republic. Previously, we only reported the presence of some of these viruses. However, we were unable to present their full-length sequences because of the unavailability of complete reference sequences in the public databases. Therefore, our efforts were aimed at obtaining and analysing full-length sequences of *Tymoviridae* viruses to understand the viral diversity in the Czech Republic. We obtained complete sequences of GFkV (grapevine fleck virus) and GRGV (grapevine red globe virus) from the genus *Maculavirus* and GRVFV (grapevine rupestris vein feathering virus) and GSyV-1 (grapevine Syrah virus 1) from the genus *Marafivirus*. Mixed infections with these viruses were observed, as well as several variants of these viruses in the same plant and we found evidence of intraspecific recombination in our Czech GRVFV, GSyV-1 and GRGV isolates. A high divergence in ORF3 and ORF4 at the amino acid level was observed between the Czech and Italian GFkV isolates. Grapevine asteroid mosaic-associated virus (GAMaV) was found for the first time in the Czech Republic.

viruses

MDPI

Article

Multiple Infections with Viruses of the Family *Tymoviridae* in Czech Grapevines

Marcela Komínková [1,2], Karima Ben Mansour [1,2], Petr Komínek [1,*], Jana Brožová [1] and Radomíra Střalková [3]

[1] Ecology, Diagnostics and Genetic Resources of Agriculturally Important Viruses, Fungi and Phytoplasmas, Crop Research Institute, Drnovská 507, 161 06 Prague, Czech Republic; kominkova@vurv.cz (M.K.); karima.ben_mansour@vurv.cz or karina79@hotmail.fr (K.B.M.); brozova@vurv.cz (J.B.)
[2] Department of Plant Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Kamýcká 129, 165 00 Prague, Czech Republic
[3] Crop Research Institute, Prague, Research Station for Viticulture Karlštejn, Karlštejn 98, 267 18 Karlštejn, Czech Republic; stralkova@vurv.cz
* Correspondence: kominek@vurv.cz

Abstract: This study focused on the viruses of the *Tymoviridae* family that infect grapevines in the Czech Republic. Complete sequences of GFkV (grapevine fleck virus) and GRGV (grapevine red globe virus) from the genus *Maculavirus* and GRVFV (grapevine rupestris vein feathering virus) and GSyV-1 (grapevine Syrah virus 1) from the genus *Marafivirus* were obtained using high-throughput sequencing of small RNAs and total RNAs. Mixed infections with these viruses were observed, as well as several variants of these viruses in the same plant. Phylogenetic analysis showed the position of the newly obtained virus isolates within the *Tymoviridae* family. Recombinant analysis provided evidence of single and multiple intraspecific recombinations in GRGV, GSyV-1, and GRVFV. Additionally, GAMaV, a grapevine virus from the genus *Marafivirus*, was reported for the first time in the Czech Republic.

Keywords: high-throughput sequencing; phylogenetic analysis; grapevine fleck virus; grapevine rupestris vein feathering virus; grapevine red globe virus; grapevine syrah virus 1; grapevine asteroid mosaic-associated virus

Citation: Komínková, M.; Ben Mansour, K.; Komínek, P.; Brožová, J.; Střalková, R. Multiple Infections with Viruses of the Family *Tymoviridae* in Czech Grapevines. *Viruses* **2024**, *16*, 343. https://doi.org/10.3390/v16030343

Academic Editors: Renyi Liu and Svetlana Y. Folimonova

Received: 11 January 2024
Revised: 7 February 2024
Accepted: 21 February 2024
Published: 23 February 2024

1. Introduction

Grapevine (*Vitis vinifera* L.) is a well-established cultivated perennial crop of great economic importance worldwide. Due to its vegetative propagation and multiplication, more than 100 viruses of different taxonomic groups have accumulated in grapevine genotypes during thousands of years of cultivation [1]. Many reports on the grapevine virome have been published worldwide [2–6] and including reports from the Czech Republic [7]. Thus, many viruses from different families infecting grapevines, including viruses from the *Betaflexiviridae* family, have been reported in the Czech Republic. These include grapevine pinot gris virus (GPGV), a member of the *Trichovirus* genus [8], as well as grapevine virus A (GVA) and grapevine virus B (GVB), members of the *Vitivirus* genus. Additionally, grapevine rupestris stem pitting-associated virus-1 (GRSPaV-1), a *Foveavirus*, has also been reported. From the family *Secoviridae*, grapevine fanleaf virus (GFLV), and Arabis mosaic virus (ArMV), both members of the *Nepovirus* genus, and from the family *Closteroviridae*, and grapevine leafroll-associated virus 1 (GLRaV-1) and grapevine leafroll-associated virus 3 (GLRaV-3), members of the *Ampelovirus* genus, have been reported in the Czech Republic [9]. In the present study, we specifically focused on grapevine viruses belonging to the *Tymoviridae* family. Two genera of the *Tymoviridae* family contain viruses that infect grapevines: *Maculavirus* and *Marafivirus*.

Two viruses belonging to the *Maculavirus* genus have been identified in the Czech Republic: grapevine fleck virus (GFkV), which belongs to the *Maculavirus vitis* species

(the binomial nomenclature used for virus taxonomy by the International Committee on Taxonomy of Viruses [10,11]), has a worldwide distribution [12–14] and was first reported in the Czech Republic by serological means [15]. The virus is phloem restricted [16]. In young leaves of *Vitis rupestris* Scheele, the virus causes the clearing of the veinlets, which are stripped of color. The virus is latent in *Vitis vinifera* L. [17]. It does not have a known vector but can be transmitted via grafting and dispersed by the exchange of grapevine propagation materials. The second maculavirus is the grapevine red globe virus (GRGV), which also induces specific symptoms in *Vitis rupestris* Scheele but does not cause any symptoms in *Vitis vinifera* L. [18,19]. The virus was first reported in grapevines from Italy and Albania [18] and has been documented in many other European [5,20–25] and non-European countries (USA [20], China [26], Brazil [6], Iran [27], Japan [28], and Australia [29]). GRGV was first detected in Czech siRNA data during a comparative study on the efficiency of different bioinformatic pipelines in 2019. Only a small genomic fragment was obtained from this study due to the limited availability of complete GRGV sequences in the NCBI database [30].

The *Marafivirus* genus of the *Tymoviridae* family has two viruses: Grapevine rupestris vein feathering virus (GRVFV), which was first reported in Greece [20] and has since been identified in many grapevine-growing regions worldwide [3,31]. In 2016, it was first reported in the Czech Republic through HTS and confirmed through the use of RT-PCR [7]. The virus causes mild asteroid symptoms in *Vitis vinifera* L. and vein feathering in *Vitis rupestris* Scheele [20]. The second marafivirus is the grapevine syrah virus 1 (GSyV-1), from the species *Marafivirus syrahensis*, first reported in the USA on *Vitis vinifera* cv Syrah [32]. Since then, it has been found in Chile [33], Italy [34], Hungary [35], South Africa [36], China [37], Croatia [38], Spain [39], Korea [40], Russia [41], and many other countries. The virus was reported in the Czech Republic and Slovakia after an investigation using an improved RT-PCR protocol on grapevines collected from both countries [42]. There is limited data on the effect of GSyV-1 on grapevine production [3,38]. GSyV-1 is not limited to infecting *Vitis vinifera*, as it has also been found in wild blackberries under the name Grapevine virus Q [43].

Previously, we only reported the presence of these viruses in the Czech Republic. However, we were unable to present their full-length sequences because of the unavailability of complete reference sequences in the public databases. This is no longer the case, thanks to the widespread use of HTS for plant viruses and, in particular, for grapevine virus genomics [2,4,44], which has resulted in the presence of a large and increasing number of full-genome sequences of grapevine viruses. Therefore, our efforts were aimed at obtaining and analyzing full-length sequences of *Tymoviridae* viruses to understand the viral diversity in the Czech Republic.

2. Materials and Methods

2.1. Sample Collection and RT-PCR

Grapevine plants for the present work were obtained from the collection of plant viruses at the Crop Research Institute, Prague (collection VURV-V, part of the collection VURV, officially recognized by the World Federation for Culture Collections), from the basic and prebasic propagation material vineyards at the Research Station for Viticulture Karlštejn (which is a part of the Crop Research Institute, Prague), and from surveys conducted in the Czech viticultural regions [7,9,15]. Among the hundreds of grapevines surveyed and tested, twelve grapevines were selected for the present study based on the positive RT-PCR reaction with generic primers for the family *Tymoviridae* [18] (Table 1). None of the selected grapevines exhibited any symptoms of viral infection.

Table 1. Grapevine plants used for the analysis of viruses from the family *Tymoviridae*.

Grapevine Plant	Cultivar	Origin	Locality
TI23	Rootstock Kober 125AA	Collection of plant viruses VURV-V	Prague–Ruzyně, district Prague-city
KA1	Müller-Thurgau, clone MT25/7	Prebasic propagation material	Karlštejn, district Beroun
KA3	Müller-Thurgau, clone MT30/34	Prebasic propagation material	Karlštejn, district Beroun
KA7	Müller-Thurgau, clone MT25/7	Prebasic propagation material	Karlštejn, district Beroun
KA8	Traminer	Basic propagation material	Karlštejn, district Beroun
LAM3	Blauer Portugieser	Vineyard survey	Lampelberg, Vrbovec, district Znojmo
LAM8	Grüner Veltliner	Vineyard survey	Lampelberg, Vrbovec, district Znojmo
BLA1	Riesling	Vineyard survey	Blatnice, district Hodonín
BLA2	Grüner Veltliner	Vineyard survey	Blatnice, district Hodonín
TVR11	Blaufränkisch	Vineyard survey	Tvrdonice, district Břeclav
LAN21	Müller-Thurgau	Vineyard survey	Lanžhot, district Břeclav
LUZ5	Sauvignon	Vineyard survey	Lužice, district Hodonín

Total RNA was isolated from grapevine leaves using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA). The quality and quantity of the RNA was assessed using agarose gel electrophoresis and photometric analysis (NanoDrop; Thermo Fisher, Waltham, MA, USA). It was then subjected to RT-PCR using a One-Step RT-PCR Kit (Qiagen GmbH, Hilden, Germany) and generic primers [18]. The presence of amplicons (386 bp) indicates the presence of either maculavirus or marafivirus in the grapevines studied. The PCR products were excised from the agarose gel and purified using the Min Elute ™ Gel Extraction Kit (Qiagen GmbH, Hilden, Germany). The purified PCR products were cloned into the pGEM-T Easy plasmid (Promega, Madison, WI, USA). A minimum of ten clones were taken from each sample and sequenced commercially (Macrogen Europe, Amsterdam, The Netherlands).

2.2. High-Throughput Sequencing (HTS)

Based on the sequencing results of the PCR products, four grapevines (TI23, LAM3, LAM8, and BLA1) were selected for high-throughput sequencing (HTS). Ribosomal RNAs were removed from total RNA preparations of the grapevine plants using a RiboMinus Plant Kit for RNA-Seq (ThermoFisher Scientific) according to the manufacturer's instructions. Total RNA libraries were then prepared using the TrueSeq Stranded mRNA Kit (Illumina), following its simplified protocol without poly-A RNA enrichment. They were sequenced commercially using a MiSeq instrument (Illumina) at the BIOCEV Center in Vestec, Czech Republic, for paired read sequencing (2 × 150 bp). At first, the four grapevines were sequenced together in a single MiSeq run. To obtain a higher sequencing depth, the sample BLA1 was run again as a single MiSeq run.

The total RNA from TI23 was extracted from 1 g of scraped phloem, and libraries of small RNAs were prepared and sequenced using Illumina HiScanSQ (SELGE, University of Aldo Moro, Bari, Italy) as described in [7].

2.3. Bioinformatic Analysis

The data obtained from the two MiSeq runs were analyzed using Geneious Prime software, version 2022.1.1. The resulting reads were trimmed using the BBduk trimmer embedded in Geneious. Duplicate reads were removed, and the remaining reads were paired and merged. De novo assembly was performed using Geneious assembler with default parameters and a sensitivity set to medium/fast. De novo contigs were then annotated using the BLAST module of Geneious to identify homology to a local database of viruses and viroids downloaded from the NCBI RefSeq database (https://www.ncbi.nlm.nih.gov/refseq/ accessed on 20 March 2023).

Contigs identified as members of the *Tymoviridae* family were checked using the NCBI's online BLAST tool to find the closest full-length sequences based on E-value. Matching GenBank sequences were then used to map the HTS reads. Mapping was performed using the Geneious mapper with the following parameters: high sensitivity/medium and a minimum mapping quality of 20. Supplementary Table S1 contains a list of the used reference sequences.

The dataset from the TI23 plant obtained through siRNA sequencing [7] was used to map and assemble GRGV, GRVFV, and GSyV-1 genomes. In addition to the Geneious mapper, sequencing data were uploaded to the Galaxy web platform using the public server at usegalaxy.org [45]. The BWA tool was used to map the siRNA reads using the default parameters. For the mapping and assembly of siRNA reads for a given virus, we employed only unused reads from the mapping of a second sequence of the same virus from the same plant/dataset.

Newly acquired full-length sequences were analyzed for their genomic structure. ORFs were identified using NCBI's ORF Finder online tool (https://www.ncbi.nlm.nih.gov/orffinder/ accessed on 1 November 2023). Conserved domains were identified using the NCBI Conserved Domains Search online tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi accessed on 1 November 2023).

The newly obtained complete sequences were also verified through Sanger sequencing of PCR products obtained with primers specific to the respective virus variant (Table S5).

2.4. Recombination, Phylogenetic, and Sequence Demarcation Analyses

This report examines five viruses: GFkV, GRGV, GRVFV, GSyV-1, and GAMaV. For the recombination, phylogenetic, and sequence demarcation analyses, five datasets were created. Each dataset included the Czech isolates, the subject of our study, and the remaining isolates were retrieved from the NCBI. Therefore, each of these datasets contained a total of 42 (GFkV), 19 (GRGV), 58 (GRVFV), 30 (GSyV-1), and 10 (GAMaV) sequences (Table S3).

All datasets were aligned using the MAFFT online service [46], and then trimmed to the appropriate region, to the complete coding region for GRGV, GRVFV, and GSyV-1 and to the partial replicase for GFkV, and GAMaV, due to the unavailability of complete sequences in the NCBI database, using BioEdit 7.2.5 software [47].

RDP4 software was used to screen for any potential recombination events within the different datasets. To be considered real, the recombination event must be detected by at least four algorithms with p-values < 10^{-6} across the seven algorithms implemented within this software [48,49].

The MEGA X program was used to detect the best-fitted substitution model. Subsequently, the maximum-likelihood phylogenetic trees were constructed based on either the complete genome or a partial fragment of RdRp. Their phylogenies were tested with 500 bootstrap replicates. Four outliers, (KX171167, GRGV), (MN879754, citrus virus C), (MZ440710, GSyV-1), and (MZ451101, GRVFV), representing the closely related sequence

to the virus of interest, were used as outliers to root the phylogenetic tree of GFkV, GRGV, GRVFV, and GSyV-1, respectively.

The Sequence Demarcation Tool (SDTv1.2) and BioEdit software were used to determine the pairwise identity and sequence similarities between the Czech isolates and other NCBI-retrieved sequences within each dataset [47,50].

3. Results

3.1. Sequencing of RT-PCR Products Obtained with Generic Primers for the Family Tymoviridae

A total of 120 fragments of the polymerase gene, originating from 12 Czech grapevines, were obtained through sequencing of the cloned RT-PCR products using generic primers for the *Tymoviridae* family. Among them, 43 sequences were unique, and five viruses belonging to the *Tymoviridae* family were identified. All tested grapevines were infected by at least one member of the *Tymoviridae* family, with GFkV being the most common virus (found in eight grapevines) and the most abundant virus in grapevine based on the number of sequenced clones per plant identified as GFkV, followed by GRVFV, which was found in seven grapevines. GRGV was detected in three grapevines, GSyV-1 was detected in two grapevines, and grapevine asteroid mosaic-associated virus (GaMaV) was detected in one grapevine (KA7). Five grapevines were infected with only one member of the family *Tymoviridae*. In addition, seven grapevines were found to be co-infected with multiple viruses (Table 2). Unique sequences obtained from the polymerase gene of *Tymoviridae* members were submitted to GenBank and are available under acc. nos. OR826216-OR826260.

Table 2. Identification and counting of viral clones from the family *Tymoviridae* using RT-PCR with generic primers.

Plant\Virus	GFkV	GRGV	GRVFV	GSyV-1	GAMaV
TI23		2	6	2	
KA1	8		2		
KA3	4		6		
KA7	8				2
KA8		10			
LAM3	5		5		
LAM8	2		8		
BLA1		5	2	3	
BLA2			10		
TVR11	10				
LAN21	10				
LUZ5	10				

3.2. HTS Data Analysis and Identification of Viruses

Grapevines with interesting characteristics were selected for subsequent HTS analyses. Based on the sequencing results using generic primers, two plants (LAM3 and LAM8) were found to be infected with GFkV and GRVFV. GFkV was the virus of interest in these plants, as it has only one complete sequence present in databases worldwide, without any ambiguities. Plant TI23 was selected as a positive control (internal standard for this study) because it was previously identified as a grapevine infected with several members of the *Tymoviridae* family. Additionally, plant BLA1 was selected because it contains three members of the family *Tymoviridae*, namely GRVFV, GRGV, and GSyV-1.

The MiSeq run with four total RNA libraries produced between 186,218 and 908,759 unique reads per library, resulting in the identification of seven viruses (GLRaV-1, GRSPaV-1, GVA, GVB, GPGV, GFkV, and GSyV-1) and two viroids (Hop stunt viroid (HSVd) and grapevine

yellow speckle viroid 1 (GYSVd-1)). Although all four plants contained a member of the *Tymoviridae* family as detected through RT-PCR using generic primers, the HTS run allowed for the detection of a *Tymoviridae* family member in only three of the plants: BLA1 contained a mapped read of GSyV-1 and plants LAM3 and LAM8 had GFkV reads (422 and 296 reads, respectively). The library from grapevine TI23 contained 277,961 unique reads, allowing for the identification of only non-*Tymoviridae* viruses. None of the reads from this plant/library were mapped to any member of the *Tymoviridae* family, even after attempts with different parameters of the mapping algorithm or using the newly obtained *Tymoviridae* genomes as references (Table 3).

Table 3. Results of the HTS MiSeq run with multiple libraries of total RNA.

Plant	Unique Reads in Total	Virus/Viroid Detected	Unique Reads Mapped to Virus/Viroid	Genome Coverage (%)
TI23	277,961	GVA	188	41.5
		GVB	38	23.7
		GRSPaV-1	161	61.7
		GLRaV-1	1357	98.8
		HSVd	36	100
LAM3	908,759	GFkV	422	92.8
		GRSPaV-1	952	99.99
		GYSVd-1	62	100
		HSVd	75	100
LAM8	328,278	GFkV	422	92.8
		GRSPaV-1	1135	99.9
		GYSVd-1	34	100
		HSVd	46	100
BLA1	186,218	GSyV-1	1	2.2
		GPGV	21	16.8
		GRSPaV-1	58	68.5
		GYSVd-1	66	100
		HSVd	54	100

Having obtained this result, we focused on a library from the plant BLA1 and reran it as a single MiSeq run. This approach gave us 9,590,966 unique reads, allowing for higher coverage to reference sequences and the detection of three members of the *Tymoviridae* family in this single library/plant—GRGV, GRVFV, and GSyV-1. However, the genome coverage was insufficient to construct the complete genome of any of the three *Tymoviridae* members (Table 4).

Finally, we used an older HTS dataset of siRNA from the TI23 plant. It contains 19,854,724 unique reads. The presence of two tymoviridae viruses (GRVFV and GSyV-1) was reported and confirmed in a previous study [7]. Moreover, GRGV was later confirmed in the data [30]. In this work, the three previously identified viruses of the *Tymoviridae* family—GRVFV, GRGV, and GSyV-1—were successfully mapped into different reference sequences. All three viruses were present in the plant, and two genomic sequences of each virus were obtained (Table 5).

Table 4. Results of the HTS MiSeq run with a single library from total RNA. The viruses highlighted in bold are members of the *Tymoviridae* family.

Plant	Unique Reads in Total	Virus/Viroid Detected	Unique Reads Assigned to Virus	Genome Coverage (%)
BLA1	9,590,966	**GRGV**	41	29.8
		GRVFV	9	9.4
		GSyV-1	47	32.3
		GPGV	265	81.8
		GRSPaV-1	1705	99.7
		GYSVd-1	293	100
		HSVd	670	100

Table 5. Results of mapping and obtaining full-length sequences of viruses from the family *Tymoviridae* from the siRNA library of the TI23 plant, with 19,854,724 total unique reads.

Virus Isolate	GenBank Acc. No.	Sequence Length	Reference Sequence for Mapping	Reference Length	Pairwise Nucleotide Identity with Reference Sequence (%)	Coverage of Reference Sequence (%)	No. of Reads Mapped to the Reference Sequence Using Galaxy	No. of Reads Mapped to the Reference Sequence Using Geneious
GRGV-1	OR787584	6849	MZ451070	6850	96.92	100.00	287,329	502,284
GRGV-2	OR787585	6837	KX171167	6851	97.18	100.00	352,316	413,327
GSyV-1-3	OR787586	6481	FJ436028	6506	87.09	100.00	392,806	209,449
GSyV-1-4	OR787587	6482	KP221255	6482	96.7	100.00	383,796	311,300
GRVFV-5	OR787588	6588	MZ451085	6718	85.68	100.00	539,512	995,741
GRVFV-6	OR787589	6727	MT084814	6718	96.53	100.00	400,088	477,244

The use of two different platforms (Geneious and Galaxy) for the mapping of viruses of the family *Tymoviridae* gave slightly different results for a number of mapped reads, except for GRGV-1 and GRVFV-5, where the number of mapped reads for Geneious was almost double that for the Galaxy platform, and GSyV-1-3, where the number of mapped reads for the Galaxy platform was approximately twice that for Geneious. However, the consensus sequences from both platforms were also very similar, reaching 99% identity. The Galaxy results were used to construct the final sequences. The sequences were verified via Sanger sequencing of RT-PCR products obtained with variant-specific primers (Table S5).

3.3. Grapevine Fleck Virus (GFkV)

The virus was found in eight plants via polymerase region sequencing (Table 1). Using HTS, the virus was found only in LAM3 and LAM8 plants. The reads that were mapped to the GFkV reference sequence provided 92.8% (LAM3) and 75.9% (LAM8) genome coverage (Table 4). Gaps in the HTS sequence were filled via Sanger sequencing of RT-PCR products. The complete GFkV sequence from the LAM3 library was submitted to GenBank and is available under accession number OR701334. The genomic structure of the full-length GFkV sequence is typical of maculaviruses, with one large open reading frame (ORF) that encodes a polyprotein of 1949 aa and an additional ORF that encodes a coat protein of 230 aa.

Comparison of the genome of the Czech isolate (OR701334) with the Italian NCBI sequence (NC_003347), the only full-length sequence without ambiguity, showed 91.6% sequence similarity, and the viral genomes' structure was identical. A relatively high amino acid divergence between the two sequences was observed in ORF3 (88.3%) and ORF4 (87.2%) (Table 6).

Table 6. Nucleotide and amino acid similarities (%) between Czech and Italian sequences of GFkV.

Italian NC_003347	5′UTR	Rep		CP		ORF3		ORF4		3′UTR
	nt	nt	aa	nt	aa	nt	aa	nt	aa	nt
Czech OR701334	91.7	91.4	95.8	95.5	99.1	94.4	88.3	94.3	87.2	100

An ML phylogenetic tree was constructed from the partial fragment of the replicase using the best-fitted method (HKY + G + I). The tree defines two clades: Clade I contains only two isolates (OR826257 and OR826258) of Czech origin, separated from all other members assigned to Clade II (Figure 1). They shared a pairwise sequence similarity of 98.5% according to the SDT result and between 80.1–85.7% with all of the remaining isolates, members of Clade II (Figure S1). Clade II is subdivided into two sub-clades: A and B. The remaining thirteen Czech isolates clustered into Sub-clade B with members of various origins, showing high genetic diversity; three sequences (OR826245-OR826247) isolated from the same grapevine LUZ5 and one sequence (OR826248) isolated from grapevine LAM8 clustered together, with a pairwise sequence similarity of 93% between isolates from both grapevines. Two isolates (OR826249 and OR701334) recovered from grapevines TVR11 and LAM3, respectively, had a pairwise similarity of 90.6% between them, and both clustered with isolates from the USA and Russia. Seven isolates extracted from grapevines KA7, LAN21, and KA1 (OR826250-OR826256) shared a pairwise sequence similarity varying between 94.7–99.7% and clustered with isolates from Russia and Switzerland. The SDT results showed a genetic variation of 12.6% between the Czech isolates extracted from eight grapevines (Figure 2).

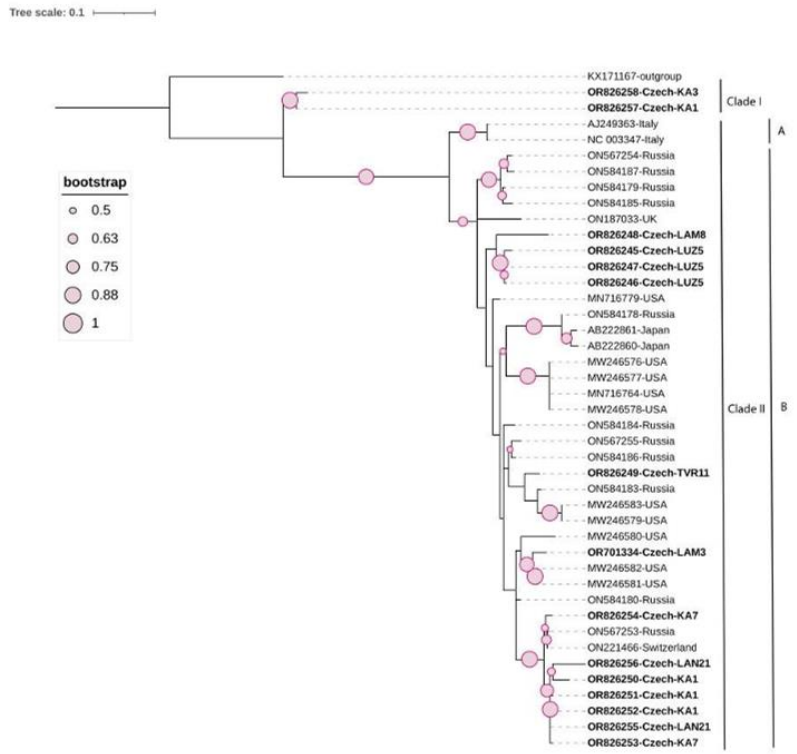

Figure 1. Maximum likelihood (ML) phylogenetic tree of 42 grapevine fleck virus isolates, including 15 Czech isolates (highlighted in bold), constructed based on the partial replicase fragment (342 nt). One sequence of GRGV (KX171167) was used as an outgroup. The tree was viewed using iTOL.

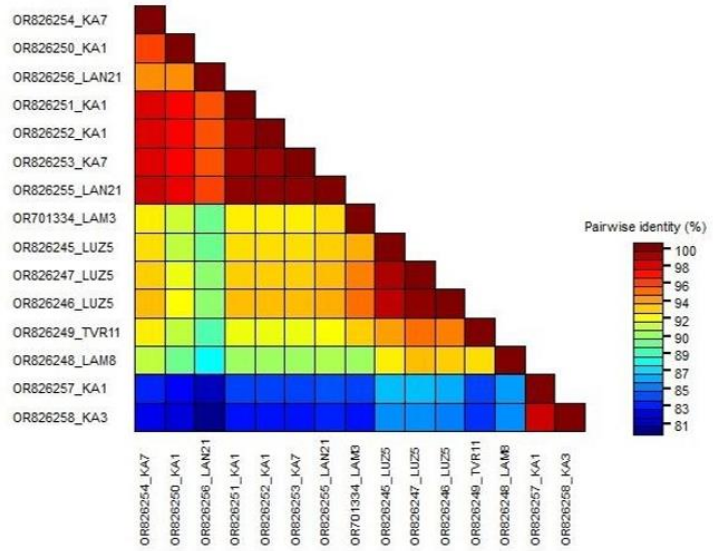

Figure 2. Matrix of pairwise nucleotide identity between 15 different Czech GFkV isolates calculated based on the partial fragment of replicase using SDTv1.2 software.

It is noteworthy that the sequence OR826257 and other isolates extracted from the same grapevine (KA1) did not cluster together; instead, the first sequence was a member of Clade I, whereas the other sequences clustered together within Clade II, in which both variants had 82.5–83.6% pairwise nucleotide identity.

3.4. Grapevine Red Globe Virus (GRGV)

Sanger sequencing of the polymerase region revealed that the virus was present in three plants (TI23, KA8, and BLA1). However, the virus was not detected in the MiSeq HTS run on multiple libraries from TI23 and BLA1, which resulted in 277,961 and 186,218 unique reads, respectively. It was only present in the MiSeq HTS run on a single library from BLA1, which resulted in a dataset containing 9,590,966 unique reads. However, the coverage of the reference sequence only reached 29.8% of its genome (Table 4). The coverage of the GRGV genome did not increase, even when the newly obtained sequences from this report were used as references for mapping.

The siRNA reads from the TI23 plant enabled the obtention of two full-length sequences: OR787584 and OR787585. They were obtained by mapping 287,329 and 352,316 reads to their respective reference sequences (Table S1). The two variants were found to be 83% identical at the nucleotide level. They contain a large ORF encoding a putative polyprotein with domains for methyltransferase, peptidase, helicase, and RNA polymerase domains. In addition, a short ORF encoding a putative coat protein is located close to the 3′ end of the genome.

Using the RDP4 program, a recombination event was identified (Table S4). It consists of a 5086 nt fragment detected in the methyltransferase, protease, helicase, RdRp, and CP. The two recombinants that share this event are (OR787585, Czech Republic) and (KX171167, Spain). The RDP4 program identified OR787584 (Czech Republic) as the major parent and MZ451074, a sequence from Canada, as the minor parent.

The ML phylogenetic tree was generated with the general time-reversible model substitution model with Gamma distribution (GTR+G) based on a complete genome (Figure 3). The tree shows two clades; Clade I includes two isolates, one of Czech origin (OR787584) and the other of Canadian origin (MZ451070); they share 97% pairwise similarity (Figure S1). Clade II consists of two sub-clades, A and B. The second sequence (OR787585) isolated from the same plant (TI23) clustered with a Spanish isolate (KX171167)

in Sub-clade A, sharing 97.2% pairwise similarity. Both sequences are recombinants, as previously mentioned. The two molecular variants (OR787584 and OR787585) exhibited a high degree of variation (17.5%) calculated based on the complete coding region (Figure S1).

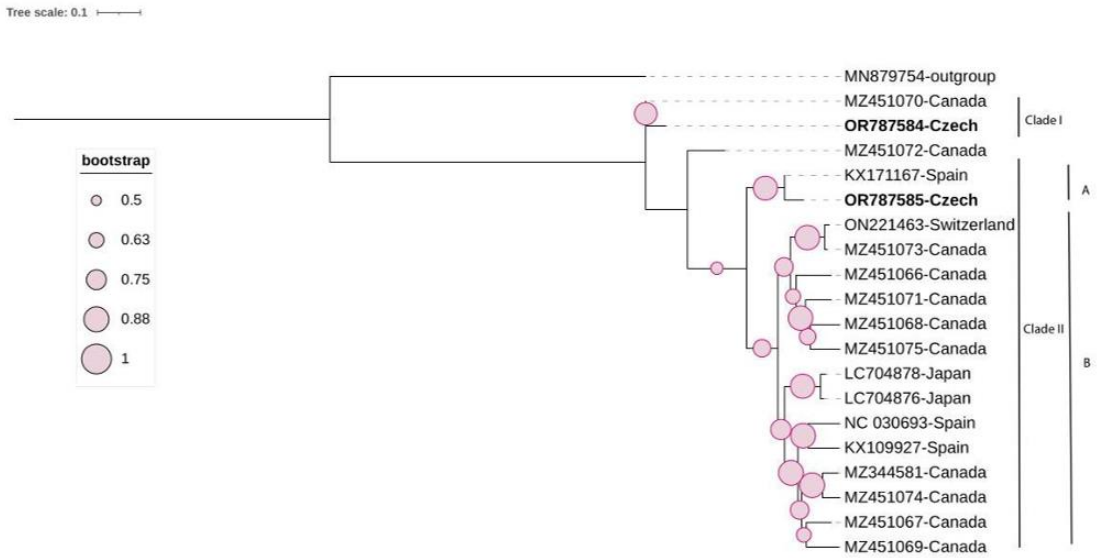

Figure 3. Maximum likelihood (ML) phylogenetic tree of 19 grapevine red globe virus isolates, including 2 Czech isolates (highlighted in bold), constructed based on their complete coding region. One sequence of Citrus virus C (MN879754) was used as an outgroup. The tree was viewed using iTOL.

3.5. Grapevine Rupestris Vein Feathering Virus (GRVFV)

Sanger sequencing of the polymerase region showed that seven out of the twelve tested plants were positive for GRVFV. Depending on the sequencing depth, the unique reads resulting from the MiSeq run on multiple libraries ranged from 186,218 to 908,759, and we could not confirm the presence of GRVFV in all of the plants (TI23, LAM3, LAM8, and BLA1). However, once the number of unique reads increased more than fifty-fold (9,590,966) following the use of an MiSeq run on a single library (BLA1), the virus could be detected. Its mapping to the reference sequence resulted in only 9.4% genomic coverage. Similarly, GRVFV was detected in grapevine (TI23) once the number of unique reads increased more than seventy-fold (19,854,724) compared to the MiSeq run on multiple libraries, allowing for 100% genomic coverage to the reference sequence. Therefore, two complete GRVFV sequences were obtained and deposited in GenBank under accession numbers OR787588 and OR787589. The sequences share 79.32% nucleotide identity. The marafiviruses share a common genome organization, which includes a large ORF that produces putative polyprotein with methyltransferase, peptidase, helicase, RNA polymerase, and coat protein domains.

The RDP4 program detected five recombination events affecting five isolates (MZ451087, MZ451088, AY706994, MN974276, and OR787588). All of the recombinant isolates, except for one isolate (OR787588) originating from the Czech Republic, were of non-European origin, (Table S4). This isolate has a 213 nt recombinant fragment detected across the replicase coding region. MZ451085 (major, Canada) and LC619667 (minor, Japan) are the detected parent for this recombination event. Based on the information available on the public database (NCBI), it appears that one or both of the predicted parents' host plants were from non-*Vitis* host species, specifically *Prunus*.

The ML phylogenetic tree was constructed using the General Time Reversible model based on the complete genome. It showed the presence of three clusters: I, II, and III. The

two Czech molecular variants of GRVF clustered differently (Figure 4). The first isolate (OR787589) belonged to Cluster II, while the second isolate (OR787588) was a member of Cluster III. The pairwise nucleotide identity between both Czech molecular variants was low, at 80.5% (Figure S2).

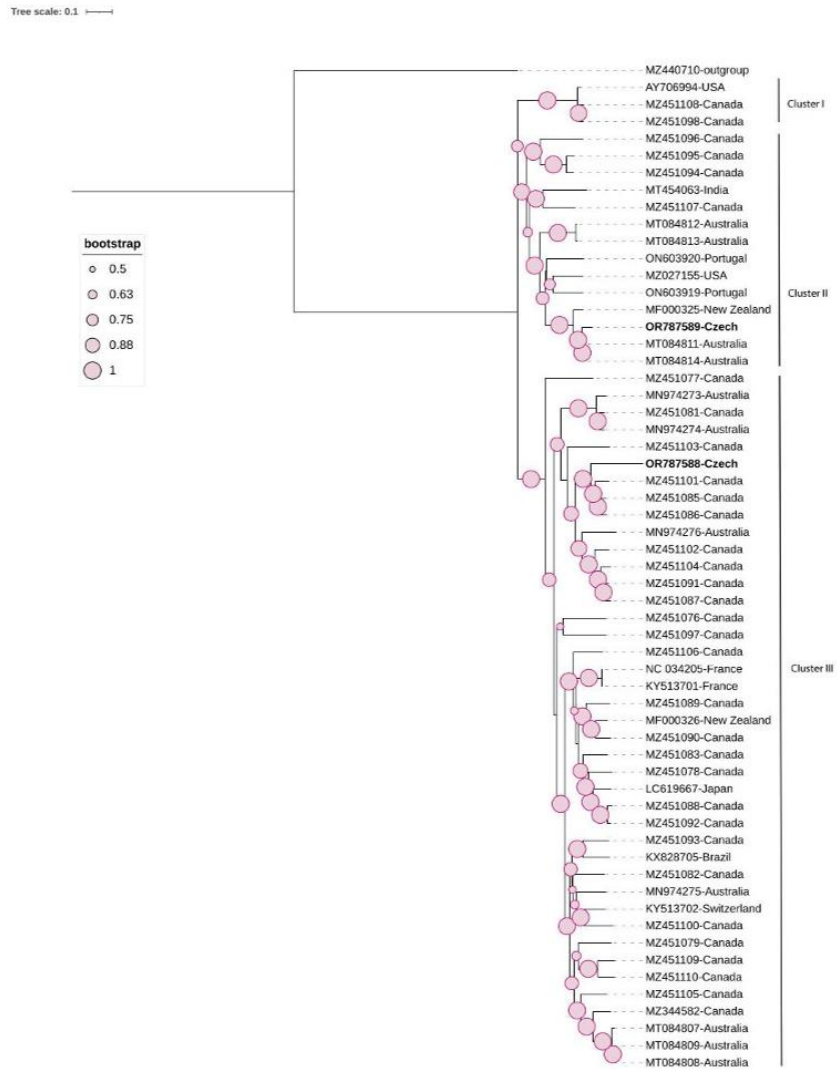

Figure 4. Maximum likelihood (ML) phylogenetic tree of 58 grapevine rupestris vein feathering virus isolates, including 2 Czech isolates (highlighted in bold), constructed based on their complete coding region. One sequence of GSyV-1 (MZ440710) was used as an outgroup. The tree was viewed using iTOL.

3.6. Grapevine Syrah Virus-1 (GSyV-1)

The MiSeq run on multiple libraries showed the presence of the virus in a grapevine, BLA1, where we detected one read. The result was further confirmed via MiSeq run on a single library and Sanger sequencing. Although the virus was detected in the TI23 grapevine through Sanger sequencing and siRNA HTS, no related reads were revealed through the MiSeq run on multiple libraries. Using the siRNA dataset, two full-length GSyV-1 sequences were obtained. Both isolates were deposited in GenBank under accession numbers

OR787586 and OR787587. The two isolates share 93% sequence nucleotide similarity and showed a genome structure typical of marafiviruses, with a large ORF encoding a 2081 aa polyprotein containing all specific domains. The ML phylogenetic tree was constructed using Hasegawa, Kishino, and Yano substitution model and Gamma distribution (HKY+G). Figure 5 shows an ML phylogenetic tree made based on the complete genome, which was subdivided into three clusters: I, II, and III. The majority of NCBI-retrieved isolates were from Canada and were mainly grouped within Clusters I and II; only one European isolate was grouped with those sequences. Cluster III was the most diverse group, containing isolates from both the European continent (the Czech Republic and Slovakia) and the American continent (USA, Canada, and Brazil). The Czech isolates formed a separate group, sharing 91.6–93% pairwise sequence identity and 81.4–90.3% pairwise sequence identity with other groups in Cluster III (Figure S3).

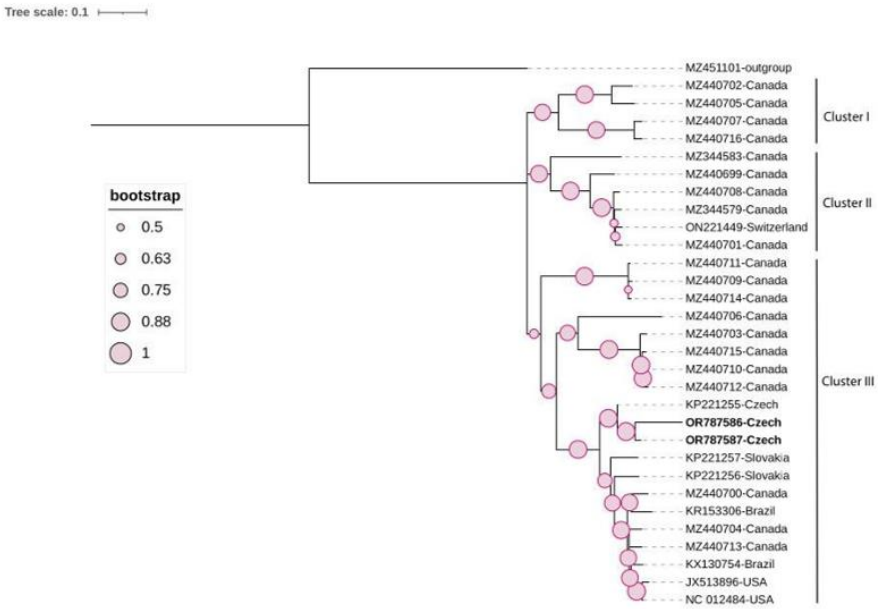

Figure 5. Maximum likelihood (ML) phylogenetic tree of 30 grapevine Syrah virus-1 isolates, including 2 Czech isolates (highlighted in bold), constructed based on their complete coding region. One sequence of GRVFV (MZ451101) was used as an outgroup. The tree was viewed using iTOL.

Recombination analysis was performed using 30 GSyV-1 sequences, revealing that seven isolates were recombinants (Table S4). The two North American isolates (JX513896, MZ440699) had only one recombination event, while five isolates from central Europe, three Czech isolates (KP221255, OR787586, and OR787587), and two Slovak isolates (KP221256-KP221257) had multiple recombination events (n = 10).

3.7. Grapevine Asteroid Mosaic-Associated Virus (GAMaV)

Sanger sequencing revealed the presence of the virus in only one plant, KA7. The partial replicase fragment was deposited in GenBank under accession number OR826259. The infected grapevine is a prebasic propagation material that originated from Viticulture Station Karlštejn. The SDT result indicates that the Czech isolate (OR826259) shares between 88.6 and 93% pairwise similarity with the other isolates retrieved from the NCBI (Figure 6). The Czech isolate shows the highest similarities with two isolates (MZ344576 and AB276378) from Canada and Japan (93%). All of the GAMaV isolates were divergent from the isolate from Switzerland.

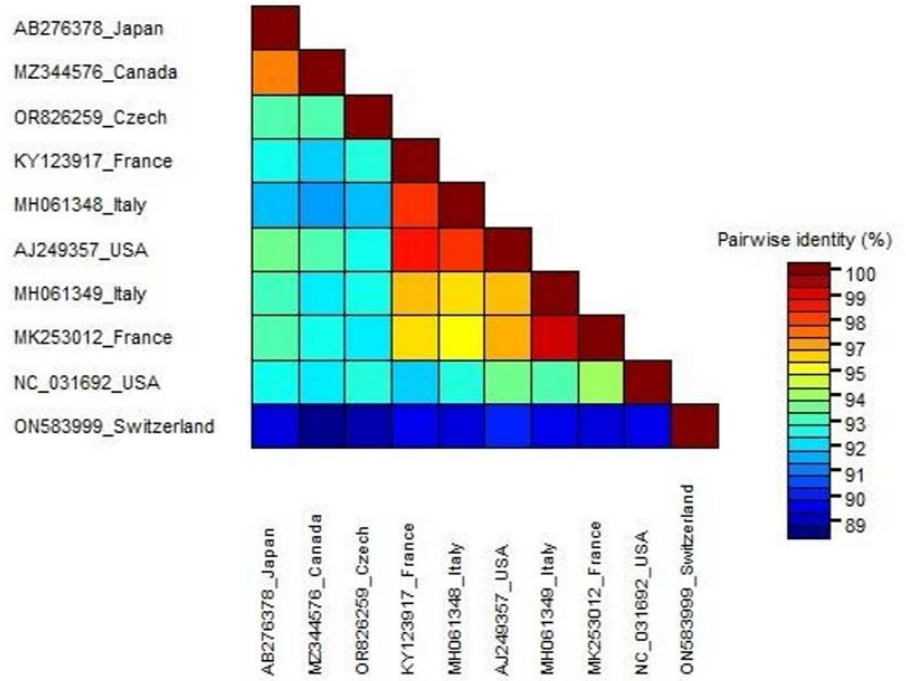

Figure 6. Matrix of pairwise nucleotide identity between 10 different GAMaV isolates calculated based on the partial fragment of replicase using SDTv1.2 software.

4. Discussion

In this study, we attempted to obtain the complete sequences of the viruses of interest using HTS. We observed that the number of viruses detected in a grapevine and the genomic coverage to the reference sequence increased with the sequencing depth. This observation was confirmed by the TI23 grapevine, which served as a positive control for the approaches using cloning of RT-PCR fragments obtained with generic primers and for the MiSeq run on multiple libraries. Thus, the failure to detect the full set of viruses that were identified and confirmed in two of our previous studies [7,30] showed that these methods could give negative results, which is in agreement with an earlier report stating that sequencing depth and viral concentrations in grapevine tissues and isolated RNAs affect the successful detection of viruses [30]. Previously, the genome coverage of GRVFV and GSyV-1 was incomplete, reaching 33% and 68%, respectively [7,30]. In this study, we applied more recent bioinformatic methods and used the newly available NCBI sequences to obtain complete genomes of members of the *Tymoviridae* family. Two platforms were used for HTS analysis: Geneious and the web-based Galaxy project. Finally, we successfully assembled one complete GFkV sequence and two complete molecular variants of GRGV, GRVFV, and GSyV-1. These viruses were found in either simple or mixed infections. Co-infection (GFkV+GRVFV) was located in four grapevines (KA1, KA3, LAM3, and LAM8), and co-infection (GFkV+GAMaV) was found in one grapevine (KA7). Two grapevines (TI23 and BLA1) were infected by three viruses (GRGV, GRVFV, and GSyV-1). Only one member of the family *Tymoviridae* was found in five grapevines. The mixed infection of GFkV and GRVFV was reported previously in Spain and Slovenia [2,51]. Another viral combination (GRGV+GRVFV) was reported previously in Spain and was found to be frequent. Based on phylogenetic and pairwise sequence analyses, a second type of mixed infection, the presence of different molecular variants of the same virus in one plant, was identified. Therefore, grapevine KA1 had two variants of GFkV, and grapevine TI23 had two molecular variants of each of the following viruses: GRGV and

GRVFV. This phenomenon has been observed in many other reports, such as the case of grapevine rupestris stem pitting-associated virus [52,53] and grapevine leafroll-associated virus 1 [54].

GFkV was the most frequently detected virus, found in eight out of the twelve examined plants. Our previous studies also reported its high occurrence in Czech grapevines [10]. According to the phylogenetic analysis based on the partial replicase, the GFkV isolates are divided into two clades, I and II, which is consistent with previous studies [16,55]. Additionally, the fifteen Czech isolates analyzed in this study were dispersed throughout the phylogenetic tree, indicating a high level of genetic divergence between them, consistent with previous research from Russia [3]. Moreover, a comparison between the Czech and Italian isolates revealed the highest divergence in amino acid identities to be 87.2% and 88.3% in ORF4 and ORF3, respectively. The effect of this divergence on the viral life cycle could not be speculated since, up to now, the role of these two proteins is still unknown [12,56]. The second most frequently detected virus via RT-PCR was GRVFV; the Czech isolates showed high genetic variability among them, which is consistent with another study that found Hungarian GRVFV isolates to be diverse [5]. Limited occurrence of three more viruses (GRGV, GSyV-1, and GAMaV) was observed. An interesting observation is that GAMaV, a marafivirus, that has been reported in California [32], Canada [57], Japan [53], Uruguay [58], France [59], Spain [60], Italy [61], Russia [3], and Hungary [5], is reported for the first time in the Czech Republic in this study, although it was found in only one plant (KA7). Expanding its genome sequence would be of great interest, especially since its detection in the Czech Republic came from an asymptomatic grapevine.

Recombination plays a crucial role in promoting adaptability to new hosts and changing environmental conditions [49,62]. It also helps to promote viral survival by reducing the number of deleterious mutations [63]. Intraspecific recombination is common in RNA viruses; few studies have screened for the presence of recombinants in viruses infecting grapevine. The most recent studies were on GPGV, which was found to be recombinant [64], GLRaV-3, GLRaV-4, GRSPaV, GVA, GVB, and GSyV-1 [6]. In this study, we looked for the possible presence of recombination and identified many recombinants. The marafiviruses (GRVFV and GSyV-1) had more recombination events than the maculavirus (GRGV). The first interesting observation was for GRVFV. All identified recombinants had either one or both parental sequences from a non-*Vitis* host (*Prunus*) based on the information communicated on the public database (NCBI). Although we are taking this information with caution, this primary result suggests that recombination may occur between GRVFV isolates from *Vitis* and non-*Vitis* hosts. An insect vector has been proposed as the reason for field transmission due to multiple infections of GRVFV+GRGV within the same plant [65]. The presence of recombinants with parental sequences from non-*Vitis* hosts suggests the likely involvement of polyphagous insects in the transmission of GRVFV. The second interesting observation is the evidence of multiple recombination events for GSyV-1 isolates from central Europe. This may indicate that this population is genetically diverse because the isolates recombine freely [66]. This is consistent with a study in Hungary showing that central European isolates are genetically diverse [5].

The examined grapevines were asymptomatic, even those that had multiple infections. This observation could be explained by the effect of genotypes such as TI23, infected by seven viruses, which is an interspecific grapevine rootstock Kober 125AA that is mostly asymptomatic when infected with viruses. However, after the transmission of viruses through grafting into some susceptible genotypes, like LN33, strong symptoms may appear [67].

Grapevine viruses are considered as a serious threat to grapevine yield and quality. As the plant is a perennial and vegetatively propagated crop, an important tool to control grapevine viruses is the certification of grapevine propagation material, as defined by law. However, this is not the case for most of the viruses of the Tymoviridae family that are the subject of our study. Under current European legislation, of all the tymoviruses, only GFkV is monitored in propagating material and only in rootstocks, not in varieties. The

absence of symptoms in the plants included in our experiment confirms the low level of risk of these viruses for grapevine cultivation, and their inclusion in certification schemes for the propagation of grapevine varieties can hardly be expected. Nevertheless, care must be taken when exchanging grapevine genetic material, as this is the most common way of spreading grapevine viruses. The ever-increasing number of viruses described on the grapevine encourages caution on the part of those responsible for the health of the grapevine and, in particular, government organizations. Even harmless viruses, under certain conditions, may combine to produce devastating effects. For example, co-infection of ArMV and GVB in warm conditions had a lethal impact within two years [15], or the co-existence of grapevine leafroll-associated virus 3 (GLRaV-3) with a susceptible variety or with GVA can affect the observed symptomatology [68,69].

5. Conclusions

In our study, we found both mixed infections of different viruses of the family *Tymoviridae* and different molecular variants of the same virus within the same grapevine. We obtained full-length sequences of four grapevine viruses from the Czech Republic (GFkV, GRGV, GRVFV, and GSyV-1), and we found evidence of intraspecific recombination in our Czech GRVFV, GSyV-1 and GRGV isolates. A high divergence in ORF3 and ORF4 at the amino acid level was observed between the Czech and Italian GFkV isolates. GAMaV was found for the first time in the Czech Republic.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/v16030343/s1. Table S1: List of reference sequences used for mapping; Table S2: List of primers used for virus detection; Table S3: List of accession numbers used for phylogenetic studies; Table S4: Recombination events detected in each dataset; Table S5: Primers for verification of newly obtained full-length sequences of viruses from the family *Tymoviridae*; Figure S1: Pairwise identity matrix, determined using SDT v. 1.2 software and based on the nucleotides of the complete coding region of 19 grapevine red globe virus, using citrus virus C (MN879754) as an outgroup; Figure S2: Pairwise identity matrix, determined using SDT v. 1.2 software and based on the nucleotides of the complete coding region of 57 grapevine rupestris vein feathering virus, using sequence of grapevine Syrah virus-1 (MZ440710) as an outgroup; Figure S3: Pairwise identity matrix, determined using SDT v. 1.2 software and based on the nucleotides of the complete coding region of 30 grapevine Syrah virus-1, using grapevine rupestris vein feathering virus (MZ451101) as an outgroup.

Author Contributions: Conceptualization, P.K. and K.B.M.; methodology, P.K., K.B.M., J.B. and R.S.; software, P.K., M.K. and K.B.M.; validation, M.K. and K.B.M.; formal analysis, K.B.M.; investigation, P.K., M.K., J.B. and R.S.; resources, P.K., J.B. and R.S.; data curation, P.K. and K.B.M.; writing—original draft preparation, P.K.; writing—review and editing, P.K. and K.B.M.; visualization, K.B.M. and M.K.; supervision, P.K.; project administration, P.K.; funding acquisition, P.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Agriculture of the Czech Republic, institutional support MZE-RO0423. Research conducted at the Research Station for Viticulture Karlštejn was supported by the Ministry of Agriculture of the Czech Republic, Subsidy Programme No. 6.2.12/MZE-62216/2022-13113 (6.2.12 grapevine).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The virus genomic sequences obtained in the present work have been deposited in the GenBank database of the National Center for Biotechnology Information (NCBI) under accession numbers OR701334, OR787584-OR787589, and OR826218-OR826260. Further data that support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fuchs, M. Grapevine Viruses: A Multitude of Diverse Species with Simple but Overall Poorly Adopted Management Solutions in the Vineyard. *J. Plant Pathol.* **2020**, *102*, 643–653. [CrossRef]
2. Miljanić, V.; Jakše, J.; Kunej, U.; Rusjan, D.; Škvarč, A.; Štajner, N. Virome Status of Preclonal Candidates of Grapevine Varieties (*Vitis vinifera* L.) from the Slovenian Wine-Growing Region Primorska as Determined by High-Throughput Sequencing. *Front. Microbiol.* **2022**, *13*, 830866. [CrossRef]
3. Shvets, D.; Porotikova, E.; Sandomirsky, K.; Vinogradova, S. Virome of Grapevine Germplasm from the Anapa Ampelographic Collection (Russia). *Viruses* **2022**, *14*, 1314. [CrossRef]
4. Vinogradova, S.; Porotikova, E.; Navrotskaya, E.; Galbacs, Z.N.; Massart, S.; Varallyay, E. The First Virome of a Russian Vineyard. *Plants* **2023**, *12*, 3292. [CrossRef]
5. Czotter, N.; Molnar, J.; Szabó, E.; Demian, E.; Kontra, L.; Baksa, I.; Szittya, G.; Kocsis, L.; Deak, T.; Bisztray, G.; et al. NGS of Virus-Derived Small RNAs as a Diagnostic Method Used to Determine Viromes of Hungarian Vineyards. *Front. Microbiol.* **2018**, *9*, 122. [CrossRef]
6. Fajardo, T.V.M.; Silva, F.N.; Eiras, M.; Nickel, O. High-Throughput Sequencing Applied for the Identification of Viruses Infecting Grapevines in Brazil and Genetic Variability Analysis. *Trop. Plant Pathol.* **2017**, *42*, 250–260. [CrossRef]
7. Eichmeier, A.; Komínková, M.; Komínek, P.; Baránek, M. Comprehensive Virus Detection Using next Generation Sequencing in Grapevine Vascular Tissues of Plants Obtained from the Wine Regions of Bohemia and Moravia (Czech Republic). *PLoS ONE* **2016**, *11*, e0167966. [CrossRef]
8. Glasa, M.; Predajňa, L.; Komínek, P.; Nagyová, A.; Candresse, T.; Olmos, A. Molecular Characterization of Divergent Grapevine Pinot Gris Virus Isolates and Their Detection in Slovak and Czech Grapevines. *Arch. Virol.* **2014**, *159*, 2103–2107. [CrossRef]
9. Komínek, P. Distribution of Grapevine Viruses in Vineyards of the Czech Republic. *J. Plant Pathol.* **2008**, *90*, 357–358.
10. Walker, P.J.; Siddell, S.G.; Lefkowitz, E.J.; Mushegian, A.R.; Adriaenssens, E.M.; Alfenas-Zerbini, P.; Davison, A.J.; Dempsey, D.M.; Dutilh, B.E.; García, M.L.; et al. Changes to Virus Taxonomy and to the International Code of Virus Classification and Nomenclature Ratified by the International Committee on Taxonomy of Viruses (2021). *Arch. Virol.* **2021**, *166*, 2633–2648. [CrossRef]
11. Zerbini, F.M.; Siddell, S.G.; Mushegian, A.R.; Walker, P.J.; Lefkowitz, E.J.; Adriaenssens, E.M.; Alfenas-Zerbini, P.; Dutilh, B.E.; García, M.L.; Junglen, S.; et al. Differentiating between Viruses and Virus Species by Writing Their Names Correctly. *Arch. Virol.* **2022**, *167*, 1231–1234. [CrossRef]
12. Sabanadzovic, S.; Abou-ghanem, N.; Martelli, G.P. Grapevine Fleck and Similar Viruses. In *Grapevine Viruses: Molecular Biology, Diagnostics and Management*; Meng, B., Martelli, G.P., Fuchs, M., Golino, D.A., Eds.; Springer: Berlin/Heidelberg, Germany, 2017; pp. 167–195, ISBN 9783319577067.
13. Jo, Y.; Song, M.K.; Choi, H.; Park, J.S.; Lee, J.W.; Cho, W.K. First Report of Grapevine Fleck Virus and Grapevine Virus E in Grapevine in Korea. *Plant Dis.* **2017**, *101*, 1069. [CrossRef]
14. Crnogorac, A.; Gašpar, M.; Davino, S.; Mandić, A.; Matić, S. First Report of Grapevine Fleck Virus in Vineyards of Bosnia and Herzegovina. *J. Plant Pathol.* **2020**, *102*, 1299. [CrossRef]
15. Komínek, P.; Holleinová, V. Evaluation of Sanitary Status of Grapevines in the Czech Republic. *Plant Soil Environ.* **2003**, *49*, 63–66. [CrossRef]
16. Shi, B.J.; Habili, N.; Symons, R.H. Nucleotide Sequence Variation in a Small Region of the Grapevine Fleck Virus Replicase Provides Evidence for Two Sequence Variants of the Virus. *Ann. Appl. Biol.* **2003**, *142*, 349–355. [CrossRef]
17. Martelli, G.P.; Sabanadzovic, S.; Ghanem-Sabanadzovic, N.A.; Saldarelli, P. Maculavirus, a New Genus of Plant Viruses. *Arch. Virol.* **2002**, *147*, 1847–1853. [CrossRef]
18. Sabanadzovic, S.; Abou-Ghanem, N.; Castellano, M.A.; Digiaro, M.; Martelli, G.P. Grapevine Fleck Virus-like Viruses in Vitis. *Arch. Virol.* **2000**, *145*, 553–565. [CrossRef]
19. Ghanem-Sabanadzovic, N.A.; Sabanadzovic, S.; Martelli, G.P. Sequence Analysis of the 3′ End of Three Grapevine Fleck Virus-like Viruses from Grapevine. *Virus Genes* **2003**, *27*, 11–16. [CrossRef]
20. El Beaino, T.; Sabanadzovic, S.; Digiaro, M.; Abou Ghanem-Sabanadzovic, N.; Rowhani, A.; Kyriakopoulou, P.E.; Martelli, G.P. Molecular Detection of Grapevine Fleck Virus-like Viruses. *Vitis* **2001**, *40*, 65–68.
21. Beuve, M.; Candresse, T.; Tannières, M.; Lemaire, O. First Report of Grapevine Redglobe Virus (GRGV) in Grapevine in France. *Plant Dis.* **2015**, *99*, 422. [CrossRef] [PubMed]
22. Cretazzo, E.; Padilla, C.V.; Velasco, L. First Report of Grapevine Red Globe Virus in Grapevine in Spain. *Plant Dis.* **2017**, *101*, 264–265. [CrossRef]
23. Ruiz-García, A.B.; Zarghani, S.N.; Okic, A.; Olmos, A.; Wetzel, T. First Report of Grapevine Red Globe Virus in Grapevine in Germany. *Plant Dis.* **2018**, *102*, 1675. [CrossRef]
24. Candresse, T.; Faure, C.; Marais, A. First Report of Grapevine Red Globe Virus (GRGV) and Grapevine Rupestris Vein Feathering Virus (GRVFV) Infecting Grapevine (*Vitis vinifera*) in Portugal. *Plant Dis.* **2022**, *107*, 974. [CrossRef]
25. Dixon, M.; Fowkes, A.; Hogan, C.; Adams, I.; McGreig, S.; Pufal, H.; Ward, R.; Harju, V.; Skelton, A.; Fox, A. First Report of Grapevine Red Globe Virus in Grapevine in the United Kingdom. *New Dis. Rep.* **2022**, *46*, e12118. [CrossRef]
26. Fan, X.D.; Dong, Y.F.; Zhang, Z.P.; Ren, F.; Hu, G.J.; Li, Z.N.; Zhou, J. First Report of Grapevine Red Globe Virus (GRGV) in Grapevines in China. *Plant Dis.* **2016**, *100*, 2340. [CrossRef]

27. Nourinejhad Zarghani, S.; Khalili, M.; Dizadji, A.; Wetzel, T. First Report of Grapevine Red Globe Virus in Grapevine in Iran. *J. Plant Pathol.* **2021**, *103*, 661. [CrossRef]
28. Yamamoto, T.; Sato, H.; Suzuki, T.; Miyazaki, A.; Kitazawa, Y.; Maejima, K.; Namba, S.; Yamaji, Y. Complete Genome Sequences of Grapevine Red Globe Virus in Japan. *Microbiol. Resour. Announc.* **2022**, *11*, 12–14. [CrossRef]
29. Wu, Q.; Habili, N.; Tyerman, S.D.; Rinaldo, A.; Little, A.; Constable, F.E. First Detection of Five Previously Unreported Grapevine Viruses in Australia. *Australas. Plant Dis. Notes* **2023**, *18*, 27. [CrossRef]
30. Massart, S.; Chiumenti, M.; De Jonghe, K.; Glover, R.; Haegeman, A.; Koloniuk, I.; Komínek, P.; Kreuze, J.; Kutnjak, D.; Lotos, L.; et al. Virus Detection by High-Throughput Sequencing of Small RNAs: Large-Scale Performance Testing of Sequence Analysis Strategies. *Phytopathology* **2019**, *109*, 488–497. [CrossRef]
31. Wu, Q.; Kehoe, M.A.; Kinoti, W.M.; Wang, C.P.; Rinaldo, A.; Tyerman, S.; Habili, N.; Constable, F.E. First Report of Grapevine Rupestris Vein Feathering Virus in Grapevine in Australia. *Plant Dis.* **2021**, *105*, 515. [CrossRef]
32. Al Rwahnih, M.; Daubert, S.; Golino, D.; Rowhani, A. Deep Sequencing Analysis of RNAs from a Grapevine Showing Syrah Decline Symptoms Reveals a Multiple Virus Infection That Includes a Novel Virus. *Virology* **2009**, *387*, 395–401. [CrossRef]
33. Engel, E.A.; Rivera, P.A.; Valenzuela, P.D.T. First Report of Grapevine Syrah Virus-1 in Chilean Grapevines. *Plant Dis.* **2010**, *94*, 633. [CrossRef]
34. Giampetruzzi, A.; Roumi, V.; Roberto, R.; Malossini, U.; Yoshikawa, N.; La Notte, P.; Terlizzi, F.; Credi, R.; Saldarelli, P. A New Grapevine Virus Discovered by Deep Sequencing of Virus- and Viroid-Derived Small RNAs in Cv Pinot Gris. *Virus Res.* **2012**, *163*, 262–268. [CrossRef]
35. Czotter, N.; Szabo, E.; Molnar, J.; Kocisz, L.; Deák, T.; Bisztray, G.; Tusnády, G.; Burgyán, J.; Várallyay, É. First Description of Grapevine Syrah Virus-1 from Grapevine in Hungary. *J. Plant Pathol.* **2015**, *97*, S74. [CrossRef]
36. Oosthuizen, K.; Coetzee, B.; Maree, H.J.; Burger, J.T. First Report of Grapevine Syrah Virus 1 in South African Grapevines. *Plant Dis.* **2016**, *100*, 1252. [CrossRef]
37. Ahmed, I.; Fan, X.D.; Zhang, Z.P.; Ren, F.; Hu, G.J.; Li, Z.N.; Khaskheli, M.I.; Dong, Y.F. First Report of Grapevine Syrah Virus-1 in Grapevines in China. *Plant Dis.* **2018**, *102*, 466. [CrossRef]
38. Vončina, D.; Al Rwahnih, M.; Rowhani, A.; Gouran, M.; Almeida, R.P.P. Viral Diversity in Autochthonous Croatian Grapevine Cultivars. *Plant Dis.* **2017**, *101*, 1230–1235. [CrossRef]
39. Ruiz-García, A.B.; Sabaté, J.; Lloria, O.; Laviña, A.; Batlle, A.; Olmos, A. First Report of Grapevine Syrah Virus-1 in Grapevine in Spain. *Plant Dis.* **2017**, *101*, 1830. [CrossRef]
40. Cho, I.S.; Yang, C.Y.; Kwon, S.J.; Yoon, J.Y.; Kim, D.H.; Choi, G.S.; Hammond, J.; Moon, J.S.; Lim, H.S. First Report of Grapevine Syrah Virus 1 Infecting Grapevines in Korea. *Plant Dis.* **2019**, *103*, 2970. [CrossRef]
41. Navrotskaya, E.; Porotikova, E.; Yurchenko, E.; Galbacs, Z.N.; Varallyay, E.; Vinogradova, S. High-Throughput Sequencing of Small Rnas for Diagnostics of Grapevine Viruses and Viroids in Russia. *Viruses* **2021**, *13*, 2432. [CrossRef]
42. Glasa, M.; Predajňa, L.; Šoltys, K.; Sabanadzovic, S.; Olmos, A. Detection and Molecular Characterisation of Grapevine Syrah Virus-1 Isolates from Central Europe. *Virus Genes* **2015**, *51*, 112–121. [CrossRef]
43. Sabanadzovic, S.; Ghanem-Sabanadzovic, N.A.; Gorbalenya, A.E. Permutation of the Active Site of Putative RNA-Dependent RNA Polymerase in a Newly Identified Species of Plant Alpha-like Virus. *Virology* **2009**, *394*, 1–7. [CrossRef]
44. Di Gaspero, G.; Radovic, S.; De Luca, E.; Spadotto, A.; Magris, G.; Falginella, L.; Cattonaro, F.; Marroni, F. Evaluation of Sensitivity and Specificity in RNA-Seq-Based Detection of Grapevine Viral Pathogens. *J. Virol. Methods* **2022**, *300*, 114383. [CrossRef]
45. Jalili, V.; Afgan, E.; Gu, Q.; Clements, D.; Blankenberg, D.; Goecks, J.; Taylor, J.; Nekrutenko, A. The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2020 Update. *Nucleic Acids Res.* **2021**, *48*, W395–W402. [CrossRef] [PubMed]
46. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT Online Service: Multiple Sequence Alignment, Interactive Sequence Choice and Visualization. *Brief. Bioinform.* **2018**, *20*, 1160–1166. [CrossRef] [PubMed]
47. Hall, T.A. BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.
48. Martin, D.P.; Murrell, B.; Golden, M.; Khoosal, A.; Muhire, B. RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes. *Virus Evol.* **2015**, *1*, vev003. [CrossRef]
49. Ben Mansour, K.; Gibbs, A.J.; Komínková, M.; Komínek, P.; Brožová, J.; Kazda, J.; Zouhar, M.; Ryšánek, P. Watermelon Mosaic Virus in the Czech Republic, Its Recent and Historical Origins. *Plant Pathol.* **2023**, *72*, 1528–1538. [CrossRef]
50. Muhire, B.M.; Varsani, A.; Martin, D.P. SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS ONE* **2014**, *9*, e108277. [CrossRef]
51. Fiore, N.; Zamorano, A.; Sánchez-Diana, N.; González, X.; Pallás, V.; Sánchez-Navarro, J. First Detection of Grapevine Rupestris Stem Pitting-Associated Virus and Grapevine Rupestris Vein Feathering Virus, and New Phylogenetic Groups for Grapevine Fleck Virus and Hop Stunt Viroid Isolates, Revealed from Grapevine Field Surveys in Spain. *Phytopathol. Mediterr.* **2016**, *55*, 225–238.
52. Glasa, M.; Predajňa, L.; Šoltys, K.; Sihelská, N.; Nagyová, A.; Wetzel, T.; Sabanadzovic, S. Analysis of Grapevine Rupestris Stem Pitting-Associated Virus in Slovakia Reveals Differences in Intra-Host Population Diversity and Naturally Occurring Recombination Events. *Plant Pathol. J.* **2017**, *33*, 34–42. [CrossRef] [PubMed]

53. Nakaune, R.; Inoue, K.; Nasu, H.; Kakogawa, K.; Nitta, H.; Imada, J.; Nakano, M. Detection of Viruses Associated with Rugose Wood in Japanese Grapevines and Analysis of Genomic Variability of Rupestris Stem Pitting-Associated Virus. *J. Gen. Plant Pathol.* **2008**, *74*, 156–163. [CrossRef]
54. Komínek, P.; Glasa, M.; Bryxiová, M. Analysis of the Molecular Variability of Grapevine Leafroll-Associated Virus 1 Reveals the Presence of Two Distinct Virus Groups and Their Mixed Occurrence. *Virus Genes* **2005**, *31*, 247–255. [CrossRef] [PubMed]
55. Glasa, M.; Predajňa, L.; Komínek, P. Grapevine Fleck Virus Isolates Split into Two Distinct Molecular Groups. *J. Phytopathol.* **2011**, *159*, 805–807. [CrossRef]
56. Sabanadzovic, S.; Abou Ghanem-Sabanadzovic, N.; Saldarelli, P.; Martelli, G.P. Complete Nucleotide Sequence and Genome Organization of Grapevine Fleck Virus. *J. Gen. Virol.* **2011**, *156*, 875–879. [CrossRef]
57. Xiao, H.; Meng, B. First Report of Grapevine Asteroid Mosaic Associated Virus and Grapevine Rupestris Vein Feathering Virus in Grapevines in Canada. *Plant Dis.* **2016**, *100*, 2175. [CrossRef]
58. Jo, Y.; Choi, H.; Kyong Cho, J.; Yoon, J.Y.; Choi, S.K.; Kyong Cho, W. In Silico Approach to Reveal Viral Populations in Grapevine Cultivar Tannat Using Transcriptome Data. *Sci. Rep.* **2015**, *5*, 15841. [CrossRef] [PubMed]
59. Candresse, T.; Faure, C.; Theil, S.; Beuve, M.; Lemaire, O.; Spilmont, A.S.; Marais, A. First Report of Grapevine Asteroid Mosaic-Associated Virus Infecting Grapevine (*Vitis vinifera*) in France. *Plant Dis.* **2017**, *101*, 1061. [CrossRef]
60. Morán, F.; Canales, C.; Olmos, A.; Ruiz-García, A. First Report of Grapevine Asteroid Mosaic Associated Virus in Grapevine in Spain. *Plant Dis.* **2020**, *105*, 517. [CrossRef]
61. Porceddu, A.; Sanna, M.; Prota, V.A.; Schianchi, N.; Mercenaro, L.; Nieddu, G.; Camiolo, S. First Report of Grapevine Asteroid Mosaic-Associated Virus Infecting Grapevines (*Vitis vinifera*) in Italy. *Plant Dis.* **2018**, *102*, 2049. [CrossRef]
62. Gibbs, A.J.; Hajizadeh, M.; Ohshima, K.; Jones, R.A.C. The Potyviruses: An Evolutionary Synthesis Is Emerging. *Viruses* **2020**, *12*, 132. [CrossRef] [PubMed]
63. Acosta-Leal, R.; Duffy, S.; Xiong, Z.; Hammond, R.W.; Elena, S.F. Advances in Plant Virus Evolution: Translating Evolutionary Insights into Better Disease Management. *Phytopathology* **2011**, *101*, 1136–1148. [CrossRef] [PubMed]
64. Hily, J.M.; Poulicard, N.; Candresse, T.; Vigne, E.; Beuve, M.; Renault, L.; Velt, A.; Spilmont, A.S.; Lemaire, O. Datamining, Genetic Diversity Analyses, and Phylogeographic Reconstructions Redefine the Worldwide Evolutionary History of Grapevine Pinot Gris Virus and Grapevine Berry Inner Necrosis Virus. *Phytobiomes J.* **2020**, *4*, 165–177. [CrossRef]
65. Cretazzo, E.; Velasco, L. High-throughput Sequencing Allowed the Completion of the Genome of Grapevine Red Globe. *Plant Pathol.* **2016**, *66*, 1202–1213. [CrossRef]
66. Zakubanskiy, A.; Mitrofanova, I.; Sheveleva, A.; Chirkov, S. Analysis of Canna Yellow Streak Virus Complete Genomes Provides Evidence of Multiple Intraspecies Recombination Events. *J. Plant Pathol.* **2018**, *100*, 575–580. [CrossRef]
67. Komínek, P.; Glasa, M.; Komínková, M. Analysis of Multiple Virus-Infected Grapevine Plant Reveals Persistence but Uneven Virus Distribution. *Acta Virol.* **2009**, *53*, 281–285. [CrossRef]
68. Wang, Y.M.; Ostendorf, B.; Pagay, V. Evaluating the Potential of High-Resolution Visible Remote Sensing to Detect Shiraz Disease in Grapevines. *Aust. J. Grape Wine Res.* **2023**, *2023*, 7376153. [CrossRef]
69. Wu, Q.; Habili, N.; Kinoti, W.M.; Tyerman, S.D.; Rinaldo, A.; Zheng, L.; Constable, F.E. A Metagenomic Investigation of the Viruses Associated with Shiraz Disease in Australia. *Viruses* **2023**, *15*, 774. [CrossRef]

IV. Study 4

-Title: Characterization of prunus necrotic ringspot virus and cherry virus A infecting myrobalan rootstock

-Brief description: Prunus necrotic ringspot virus (PNRSV) and cherry virus A (CVA) are two viruses that primarily infect plants belonging to the genus *Prunus*. Full-length sequences of these two viruses were obtained via high-throughput sequencing (HTS) from symptomatic *Prunus cerasifera* plants in the Czech Republic. The two isolates were subjected to molecular characterisation by recombination and phylogenetic analyses. Graft infections of different rootstocks were performed in order to ascertain the biological characteristics of these viruses and their corresponding symptomatology.

viruses

MDPI

Article

Characterization of *Prunus Necrotic Ringspot Virus* and *Cherry Virus* A Infecting Myrobalan Rootstock

Karima Ben Mansour [1,2], Petr Komínek [1,*], Marcela Komínková [1] and Jana Brožová [1]

[1] Ecology, Diagnostics and Genetic Resources of Agriculturally Important Viruses, Fungi and Phytoplasmas, Crop Research Institute, Drnovská 507, 161 06 Prague, Czech Republic; karina79@hotmail.fr (K.B.M.); kominkova@vurv.cz (M.K.); brozova@vurv.cz (J.B.)

[2] Department of Plant Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Kamýcká 129, 165 00 Prague, Czech Republic

* Correspondence: kominek@vurv.cz

Abstract: *Prunus necrotic ringspot virus* (PNRSV) and *cherry virus* A (CVA) are two viruses that mainly infect plants of the genus *Prunus*. Full-length sequences of these two viruses, collected in the Czech Republic from *Prunus cerasifera* plants, were obtained via HTS sequencing. Phylogenetic analyses based on the NJ method and Splitstree tools showed that the Czech PNRSV isolate (ON088600-ON088602) is a divergent isolate from other molecular groups, sharing less than 97% pairwise nucleotide identity with members of other groups. The Czech CVA isolate (ON088603) belonged to molecular subgroup III-2, clustered with isolates from non-cherry hosts, and shared the highest pairwise nucleotide identity (99.7%) with an isolate of Australian origin.

Keywords: PNRSV; CVA; phylogenetic analysis; biological assay; symptoms

Citation: Ben Mansour, K.; Komínek, P.; Komínková, M.; Brožová, J. Characterization of *Prunus Necrotic Ringspot Virus* and *Cherry Virus* A Infecting Myrobalan Rootstock. *Viruses* **2023**, *15*, 1723. https://doi.org/10.3390/v15081723

Academic Editor: Aurelie Rakotondrafara

Received: 24 July 2023
Revised: 4 August 2023
Accepted: 8 August 2023
Published: 11 August 2023

1. Introduction

Two viruses, *prunus necrotic ringspot virus* (PNRSV) and *cherry virus* A (CVA), are among the most common viruses infecting sour and sweet cherries in the Czech Republic [1].

PNRSV belongs to the genus *Ilarvirus* and infects *Prunus* spp. and ornamental plants [2]. It is a positive-sense single-stranded RNA virus. It has a segmented, tripartite genome, with RNA1 and RNA2 encoding two replicase proteins, P1 and P2, respectively, and RNA 3 encoding two other proteins: movement protein (MP) and coat protein (CP) [2]. Initially, PNRSV was divided into four different molecular groups (PV32, PV96, PE5, and CH30) [3]; then, another group appeared, the SW6 [4], and, recently, another (sixth) group was proposed, the PchMX-Azt [5]. However, compared to other molecular groups, most of the reported sequences are clustered into two major groups: PV32 and PV96 [6]. PNRSV is distributed worldwide. In the Czech Republic, it was first detected serologically in sour and sweet cherries [6]. PNRSV can be transmitted by pollen, which causes rapid virus spread in orchards [7], or by seed. These two natural modes of transmission have different efficiencies depending on the host plant species [8]. PNRSV is distributed also by infected plant-propagating material, such as budwood and rootstocks. The first symptoms of PNRSV appear one year after infection, called the acute or shock stage, but, later, plants become symptomless. However, previous studies have reported that some strains cause recurrent symptoms each year [9,10]. Infected plants show different symptoms depending on the PNRSV isolate, including mosaic, ringspot, chlorosis, leaf deformation, necrosis, shot holes, and drop-off. It also significantly affects fruit yield and quality. Infections with the virus may be latent or asymptomatic [2].

The second virus is *cherry virus* A (CVA), which belongs to the genus *Capillovirus* [11]. It is a positive single-stranded RNA with two ORFs; ORF1 encodes a replicase and coat protein, while ORF2 encodes a movement protein in a different frame [11]. The symptoms

caused by this virus are considered latent or unknown because the virus is usually found in mixed infections, which complicates the association between this virus and specific disease symptoms [12]. This virus is distributed worldwide. It was first reported in 2010 in the Czech Republic in sweet and sour cherries [13]. The virus has also been found in non-cherry hosts, such as apricot, plum, peach, and Japanese apricot [14]. CVA has been divided into five molecular groups based on its RdRp [15] and six different molecular groups based on the complete genomes [14]. Recent studies have shown that CVA is clustered into seven phylogenetic groups [16]. CVA is transmitted via grafting; however, vector transmission has not yet been reported.

In the present work, these two viruses, PNRSV and CVA, were detected in a symptomatic myrobalan rootstock BN4Kr plant using high-throughput sequencing (HTS), and their complete genomes were assembled. To molecularly characterise these two Czech isolates, we screened them for recombination events using RDP4 and constructed a phylogenetic tree based on their complete coding regions. In addition, graft infections of different rootstocks were performed to determine the biological characteristics of these viruses and their corresponding symptomatology.

2. Materials and Methods

2.1. Plant Material and ELISA

In a previous study [17], several self-rooted plants of myrobalan BN4Kr (n = 55) were grown in a small open-field trial at the Crop Research Institute of Prague (CRI Prague). These plants were used to evaluate the field resistance of this rootstock (BN4Kr), together with six other rootstocks, to natural infection by plum pox virus (PPV) [17]. These plants remained PPV free throughout the four years of evaluation. Although no PPV was recorded on these plants, some ringspot symptoms were occasionally observed on them. In order to determine the causal agent behind this observation, a serological (ELISA) test using a commercially available antibody against PNRSV (Bioreba, Reinach, Switzerland) was performed to confirm the presence of this virus in these plants.

2.2. Sample Preparation and HTS

Total RNA was isolated from leaves of PNRSV ELISA-positive plant of myrobalan BN4Kr using a Spectrum Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA). Ribosomal RNA was removed from the total RNA using the RiboMinus Plant Kit for RNA-Seq (Thermo Fisher Scientific, Waltham, MA, USA). The library for HTS was prepared using the TruSeq stranded mRNA Kit (Illumina, San Diego, CA, USA), with modifications to allow for processing of total RNA preparations according to the manufacturer's instructions. Sequencing was performed on a MiSeq instrument (Illumina) at 2 × 150 nucleotides, resulting in 1,538,208 unique reads. Bioinformatic analysis was performed using Geneious Prime version 2020.2.4, as previously described [18], resulting in the assembly of the complete genomes of PNRSV and CVA.

2.3. Genome Characterization of PNRSV and CVA

The NCBI ORFfinder was used to predict the corresponding open reading frames (ORFs) from each sequence obtained (https://www.ncbi.nlm.nih.gov/orffinder, accessed on 22 May 2023). BioEdit 7.2.5 software was used to translate multiple nucleotide sequence alignments into their corresponding amino acid sequences [19]. The ExPASY ProtParam online application (https://web.expasy.org/protparam/, accessed on 31 May 2023) was used to predict the different characteristics of the PNRSV genome (molecular weight, aliphatic index, GRAVVY, GC content, total number of negatively charged residues, and total number of positively charged residues). Post-translational modifications (PTMs) were predicted using the ScanProsite server (https://www.expasy.org/resources/scanprosite, accessed on 5 June 2023).

2.4. Biological Assay and RT-PCR

Two myrobalan rootstocks (M29C and MRS 2/5) and apricot seedlings (M-VA-1) were inoculated with grafts from the PNRSV-infected BN4Kr plant. Ten plants from each rootstock were used for inoculation, and three uninfected plants were used as negative controls. The inoculated plants were grown in a screenhouse to avoid possible contamination by unwanted viral infections, and the inoculated and negative control plants were grown for five years, during which time leaf symptoms were assessed annually. The presence of the virus in the inoculated plants was further confirmed via both DAS-ELISA and RT-PCR using primers on the basis on the sequences obtained from HTS sequencing (Table S1).

2.5. Recombination Analysis

The genomes of PNRSV and CVA are tripartite and monopartite, respectively; so, four datasets were generated (Tables S2 and S3). Each dataset contained the Czech sequence—the subject of this study and all available sequences retrieved from NCBI. Thus, the first dataset contained RNA1 sequences of PNRSV (n = 41), the second dataset contained RNA2 sequences of PNRSV (n = 38), the third dataset contained RNA3 sequences of PNRSV (n = 92), and the fourth dataset contained sequences of the complete genome of CVA (n = 124).

The RNA1 and RNA2 datasets were aligned using the MAFFT online service [20] and trimmed to their respective coding regions using BioEdit version 7.2.5 [19]. To ensure that the alignment was in frame, the RNA3 and CVA datasets were prepared as concatenated ORFs aligned using the TranslatorX online server (http://translatorx.co.uk/, accessed on 23 May 2023) [21], using the encoded amino acids as a guide.

The RDP4 program was used to search for possible recombination events in these datasets using seven algorithms implemented in this program with default settings. The recombination event detected by at least three of these with a p-value $< 10^{-6}$ was considered possible [22].

2.6. Phylogenetic and Sequence Demarcation Analyses

The three PNRSV datasets were shortlisted for phylogenetic analysis to include only the NCBI-retrieved isolates with complete genomes (all three RNAs). Recombinant isolates were also excluded. In total, there were 23 sequences for each RNA molecule. Seven additional reference sequences representing the six previously reported phylogenetic groups (PV32 = Y07568, PV96 = S78312, PE5 = L38823, CH30 = AF034994, SW6 = AF013287, and FJ546090-FJ546091 = PchMX-Azt) [3–5] were added to the RNA3 dataset (n = 30) (Table S3).

After removal of the recombinant sequences, the CVA dataset contained 105 sequences.

Phylogenetic trees were constructed based on the neighbor-joining (NJ) method using the MEGAX program [23], with a bootstrap set at 1000 replicates. A sequence demarcation tool (SDTv1.2) was used to determine pairwise nucleotide identity [24], and SplitsTree4.17.2 software was used to support the different clusters of these phylogenetic groups [25].

3. Results

3.1. Genome Organization of PNRSV and CVA

The myrobalan plant infected with PNRSV and CVA is deposited in the publicly accessible collection of plant viruses at the CRI Prague (collection acronym VURV-V, deposited under ref. no. VURV-V:46.2). The collection is accessible online via web hub www.microbes.cz (accessed on 31 July 2023). Virus isolates of PNRSV and CVA are named Ruzyne after the locality where they were found.

Full sequences resulting from HTS were deposited in GenBank under the following accession numbers: PNRSV (tripartite RNAs: ON088600, ON088601, ON088602) and CVA (ON088603).

The CVA genome has a GC content of 39.7% and consists of 7415 nt, containing two open reading frames of 7029 nt coding for both replicase and CP (2342 aa) and 1392 nt coding for MP (463 aa) in another frame.

The obtained PNRSV genome was found to consist of three RNAs, RNA1 with 3332 nt and RNA2 with 2591 nt, encoding P1 and P2 proteins of 1045 aa and 799 aa, respectively. The genome of RNA3 has 1944 nt, encoding 283 aa for MP and 224 aa for CP. The remaining characteristics of the RNAs mentioned in the Materials and Methods are listed in Table 1.

Table 1. Characteristic features of the Czech PNRSV genome.

	PNRSV			
	RNA1	**RNA2**	**RNA3**	
Accession number	ON088600	ON088601	ON088602	
Size (nt)	3332	2591	1944	
GC content (%)	45.4	41.7	46.8	
5′ and 3′ UTR (nt)	29 and 164	26 and 164	174 and 169	
Start-stop codon (position)	30–3167	27–2426	175–1026	1101–1775
Protein size amino acid (aa)	1045	799	283	224
ORFs encoded proteins	P1	P2	MP	CP
Theoretical Molecular Weight (kDa)	117.28	91.27	31.32	24.88
Aliphatic index	84.78	83.20	89.49	86.47
Theoretical isoelectric point (pI)	7.72	5.10	6.47	9.37
Grand average of hydropathicity (GRAVY)	−0.288	−0.234	−0.280	−0.277
Total number of negatively charged residues (Asp + Glu)	135	115	38	20
Total number of positively charged residues (Arg + Lys)	137	81	37	27

3.2. Recombination Analysis

Recombination analysis using the RDP4 programme showed that none of the Czech PNRSV segments were recombinant; however, seven isolates retrieved from NCBI (Q15R1N, TNpeach5, Che1, Che2, 13C257, 13C258, and 13C278) were recombinants. Two isolates obtained from NCBI showed recombination events only in their RNA1 (Q15R1N = KY883333, TNpeach5 = OL800569), two isolates showed recombination events in both RNA1 and RNA2 (Che1 = MH727235, MH727230, Che2 = MH727236, MH727231), and three isolates showed only one recombination event in their RNA3 (13C257 = MZ451054, 13C258 = MZ451055, and 13C278 = MZ451059) (Table S4).

The same result was obtained for the Czech CVA isolate, which had no recombination events. However, 19 CVA isolates from the NCBI database had at least one recombination event. About one-third had multiple recombination events. Most of the recombinant CVA isolates were obtained from cherries. The RDP4 programme showed that the Czech isolate (ON088603) was the putative major parent of the two NCBI isolates (LC422952, India, apricot) and (LC752551, Korea, sweet cherry) (Table S4).

3.3. Phylogenetic Analysis and Sequence Demarcation

Three individual phylogenetic trees were constructed from the complete coding regions of RNA1, RNA2, and RNA3 of PNRSV using the neighbour-joining method. The phylogeny was tested using the bootstrap method with 1000 replicates. The phylogenetic tree constructed on the basis of the complete coding region of RNA3 (Figure 1a) showed that the Czech PNRSV isolate (ON088602) clustered separately from the other five groups (PV32, PE5, CH30, SW6, and PchMX-Azt) in a sister clade to the PV96 phylogroup. One sequence (MZ451050) clustered within the CH30 phylogroup, whereas the remaining sequences clustered mainly within PV32 and PV96.

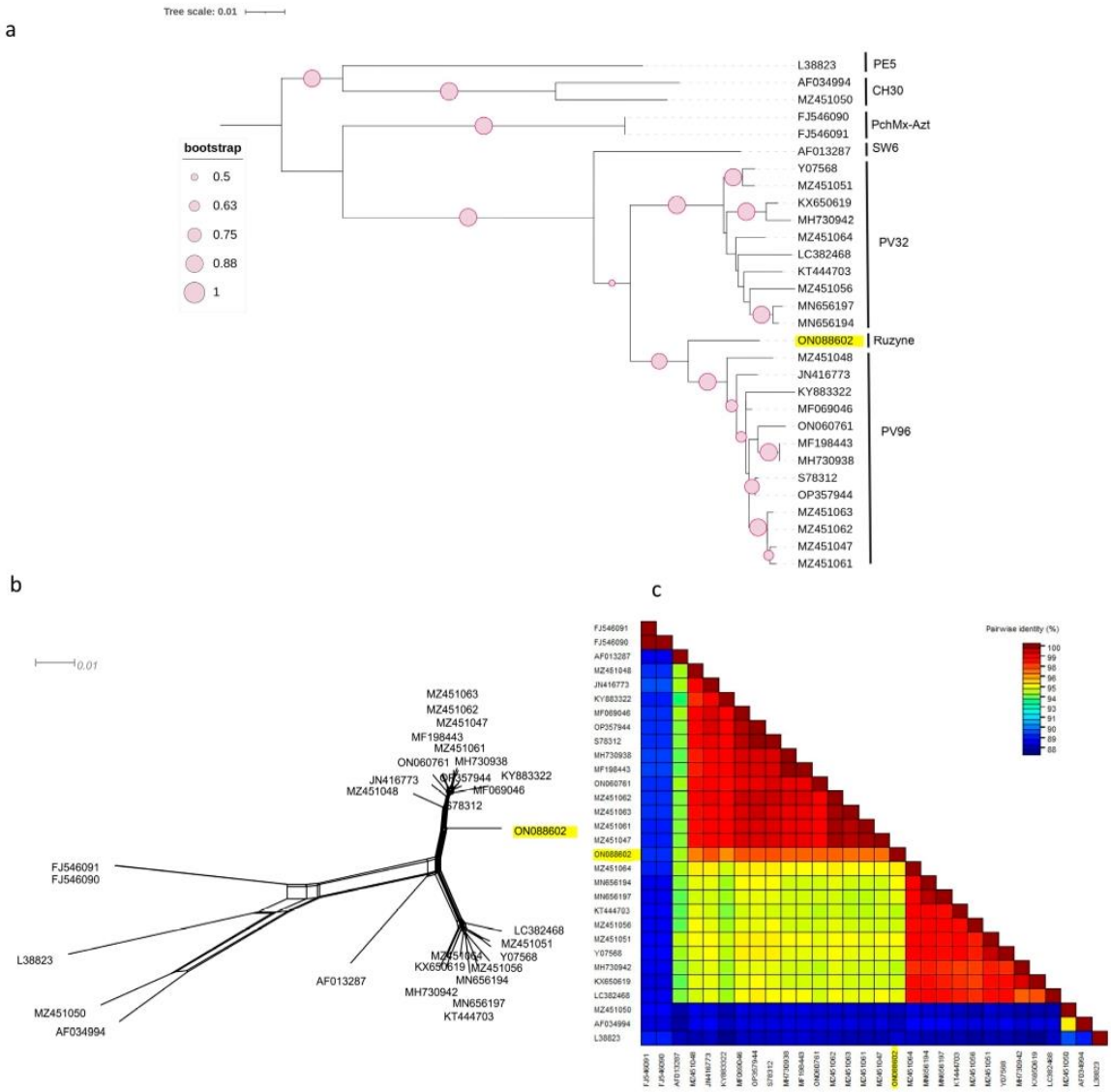

Figure 1. Characterisation of PNRSV based on the coding region of 30 RNA3 sequences using (NJ) mid-rooted phylogenetic tree (**a**), the phylogenetic network was examined using the Splitstree4.17.2. software (**b**), pairwise nucleotide identity analysis of different isolates using the SDT programme (**c**).

The same observation was made for the other two phylogenetic trees based on the complete coding regions of RNA1 and RNA2 (Figures 2a and 3a), in which the Czech isolate clustered separately from other molecular groups. Both phylogenetic trees had the same topology, with two phylogroups, I and II, in which clade II was divided into two subclades, II-1 and II-2. The Czech PNRSV isolate, the subject of this study, clustered within subclade II-1, separately from the other members.

This finding was further supported by the SplitsTree software v.4.17.2, which clearly showed that the Czech isolate was divergent from the other isolates (Figures 2b and 3b). The SDT v2.1 programme showed that the pairwise nucleotide identities of the Czech isolate in RNA1 (Figure 2c) and RNA2 (Figure 3c) ranged between 91.7–96.7% and 91.2–96.9% with members of groups I and B, respectively. For RNA3 (Figure 1c), the Czech isolate shared

pairwise nucleotide identities of 96.8–97.3%, 94.6–95.3%, and 87.7–94.4% with other PV96, PV32, and the remaining four molecular groups, respectively.

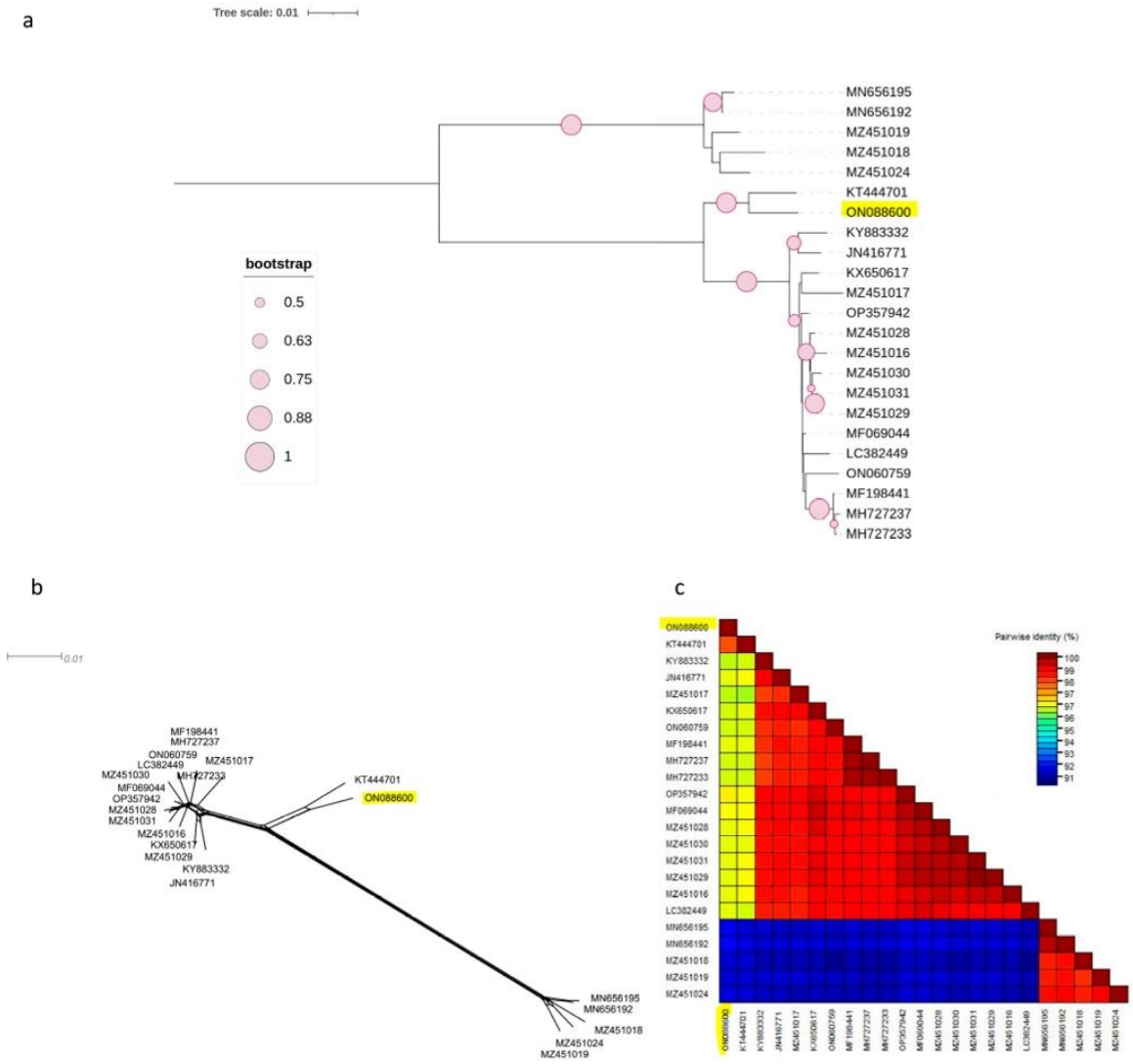

Figure 2. Characterisation of PNRSV based on the coding region of 23 RNA1 sequences using (NJ) mid-rooted phylogenetic tree (**a**), the phylogenetic network was examined using the Splitstree4.17.2. software (**b**), pairwise nucleotide identity analysis of different isolates using the SDT programme (**c**).

Phylogenetic analysis based on the complete coding region of CVA (Figure 4), using 105 non-recombinant NCBI-retrieved sequences, revealed the presence of eight distinct molecular groups (I-VIII) and three ungrouped divergent isolates (LC523006, KY510863, and KY510864). The Czech CVA isolate (ON088603) clustered in subgroup III-2 with isolates from non-cherry hosts. The Czech CVA isolate was closely related to LN879388, an Australian isolate from the same host plant, *Prunus cerasifera* Ehrh., with a pairwise nucleotide similarity of 99.7%. These observed phylogenetic groups were further investigated using SplitsTree software v.4.17.2 (Figure S1), which confirmed this clustering, and SDT software v2.1 (Figure S2), which showed that pairwise nucleotide similarities between members of each group ranged from 97.9 to 100% and that nucleotide similarities between

different groups ranged 81–87.1%. Notably, the lowest pairwise genetic similarities were observed for the divergent isolates (LC523006, KY510863, and KY510864), members of group II (KY510851 and KY510867), and group VI (MZ291923, LC523010, and KY510865), which ranged between 81.6% and 86.0% with members of other groups.

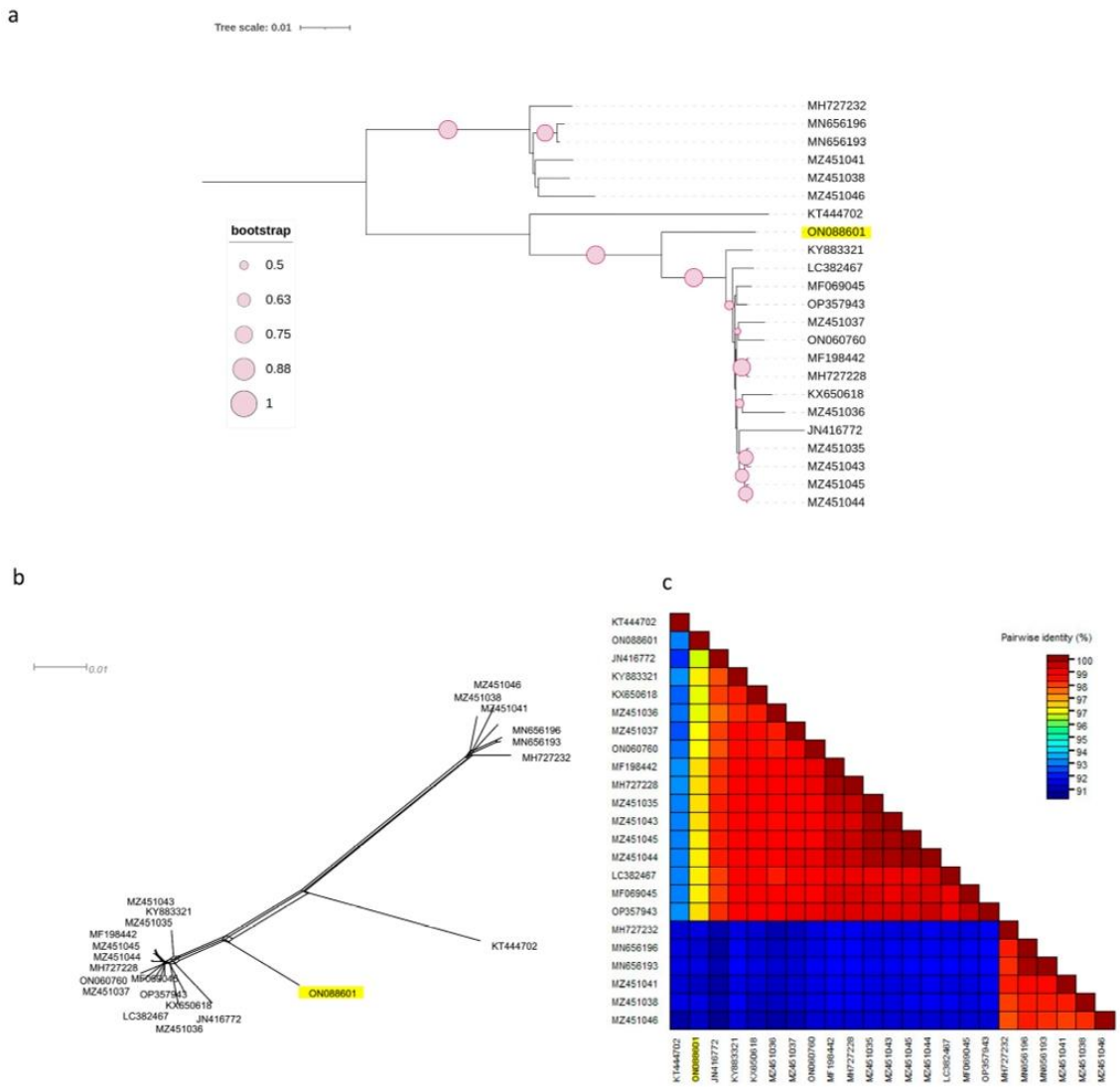

Figure 3. Characterisation of PNRSV based on the coding region of 23 RNA2 sequences using (NJ) mid-rooted phylogenetic tree (**a**), the phylogenetic network was examined using the Splitstree4.17.2. software (**b**), pairwise nucleotide identity analysis of different isolates using the SDT programme (**c**).

3.4. Variability of PNRSV Isolates

A comparison of the nucleotide sequences of the Czech PNRSV isolate (Ruzyne) with other sequences retrieved from NCBI, members of clades I and II-2, showed the presence of one SNP resulting in a non-synonymous amino acid substitution unique to the Czech isolate in coding regions of RNA1 at position 191. In addition to this unique SNP specific to the Czech isolate, it shared two SNPs with members of clade I, resulting in amino acid changes at positions 297 and 753, and 13 SNPs shared with members of clade II-2 (Table 2).

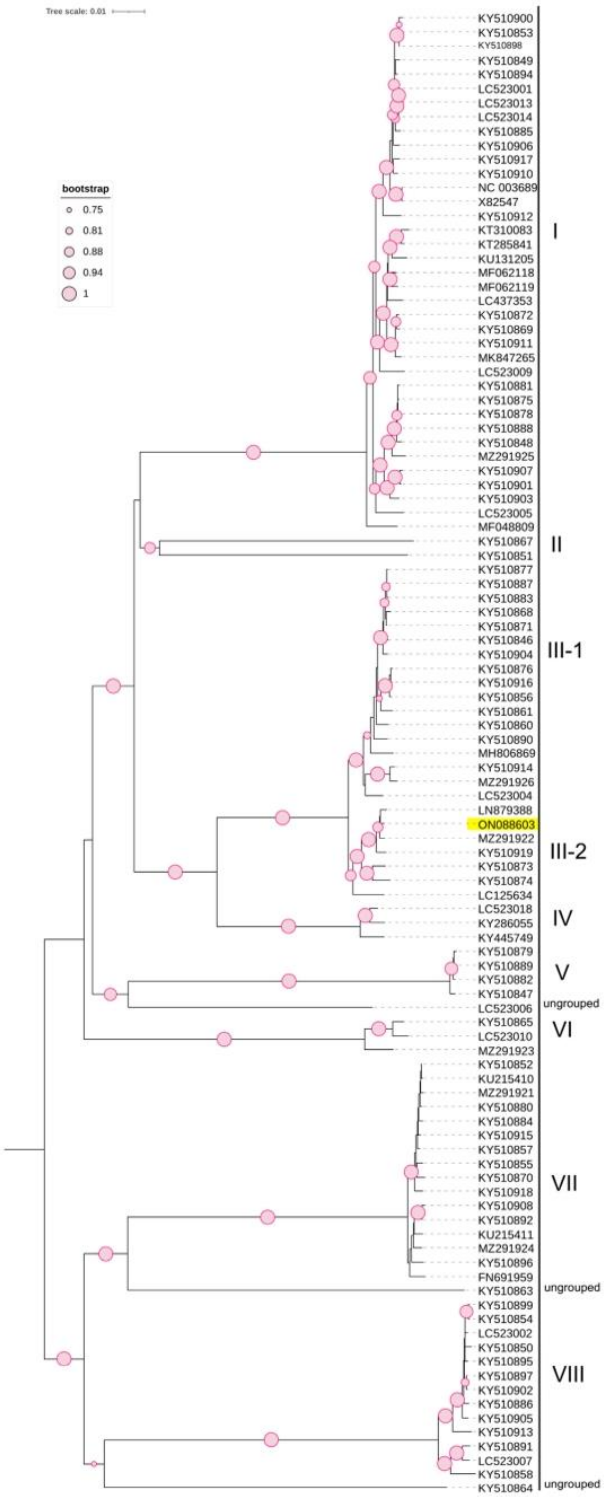

Figure 4. Phylogenetic tree based on the complete genome of nucleotide sequences of CVA constructed using the NJ method with 1000 bootstrap replicates.

Table 2. List of polymorphic sites of RNA1 and RNA2 of Czech PNRSV isolate.

Gene	Position in Genome	Position in Protein	Czech Isolate		Clade I		Clade II-2	
			Codon	AA	Codon	AA	Codon	AA
RNA1	316	106	GCG	A	TCG	S	GCG	A
	573	191	CAG	E	CAT	D	CAT	D
	890	297	AGA	R	AGA	R	AAG	K
	1285	429	TCA	S	ATA	I	TCA	S
	1291	431	CCG	P	TCG	S	CCG	P
	1304	435	GTA	V	GAA	E	GTA	V
	1306	436	GCA	A	ACT	T	GCA	A
	1349/1351	450	AGT	S	AAT	N	AGT/AGC	S
	1378/1379	460	GTT	V	ACT/ATT/	I/T	GTT	V
	1409	470	GTT	V	GCT	A	GTT	V
	1983	661	CAG	Q	CAC	H	CAG	Q
	2258	753	AGG	R	AGG	R	AAG	K
	2875	959	TCA	S	ACA	T	TCA/TCT	S
	2902	968	GGT	G	AGT	S	GGT	G
	2995	999	TTG	L	ATG	M	CTG	L
	3100	1034	GCA	A	ACA	T	GCA	A
RNA2	25	9	TCA	S	ACA	T	TCA	S
	104	35	TTT	F	TCT	S	TCT	S
	142	48	ACT	T	GCT	A	ACT	T
	179	60	ATG	M	ACG	T	ACG	T
	373	125	ATG	M	CTG	L	ATG	M
	378	126	GAA	E	GAG	E	GAC	D
	388	130	TTC	F	GTC	V	TTC/TTT	F
	409	137	GTG	V	TTG	L	GTG	V
	451	151	ATG	M	ATG	M	GTG	V
	511	171	CCT	P	TTT	F	CCT	P
	523	175	ATC	I	GTC	V	ATC	I
	553	185	GTA	V	ATA	I	GTA	V
	558	186	GAT	D	GAA/GAG	E	GAT/GAC	D
	658	220	ATT	I	GTT	V	ATT	I
	664	222	TCG	S	GTG	V	TCG	S
	826	276	ATC	I	GTT/GTC	V	ATT	I
	1550	517	AAA	K	AGA	R	AAA	K
	1706	569	CTC	L	CCT	P	CCT	P
	2303	768	AAT	N	AAT	N	AGT	S
	2329	777	GCC	A	ACC	T	GCC	A
	2380	794	TGT	C	CGT	R	TGT	C

In RNA2 there were three SNPs unique to the Czech isolate compared to other isolates, and the three resulting substitutions were located at positions 35, 60, and 569. It shared

three SNPs with members of clade I and 15 SNPs with members of clade II-2, resulting in amino acid substitutions in P2 (Table 2).

In MP, there was one unique amino acid change related to the Czech isolate, which was alanine, at position 49, compared to other isolates, all of which had isoleucine (Table 3).

Table 3. List of polymorphic sites of MP and CP genes of Czech PNRSV isolate.

Gene	Position in Genome	Position in Protein	Czech Isolate		Other Six Molecular Groups	
			Codon	AA	Codon	AA
MP	145	49	GCC	A	ATC/ATT	I
CP	406	135	GCC	A	GAC	D

In CP, the Czech isolate had an amino acid change (alanine) at position 135. In contrast, all other members of other groups had aspartic acid at this position (Table 3).

Prediction of potential PTMs revealed the presence of seven sites at positions 106, 433, 435, 436, 450, 753, and 959 in P1 and only one at position 135 in CP. These sites are putative targets for phosphorylation, N-glycosylation, and N-myristoylation.

3.5. Biological Assay

In 2019, the first symptoms of infection were observed one year after graft inoculation. Virus-inoculated plants showed different systemic symptoms, ranging from mild spots in virus-inoculated plants of myrobalan MRS 2/5 (n = 10/10) (Figure 5) to necrotic spots in leaf tissue of plants of myrobalan M29C (n = 10/10) (Figure 6).

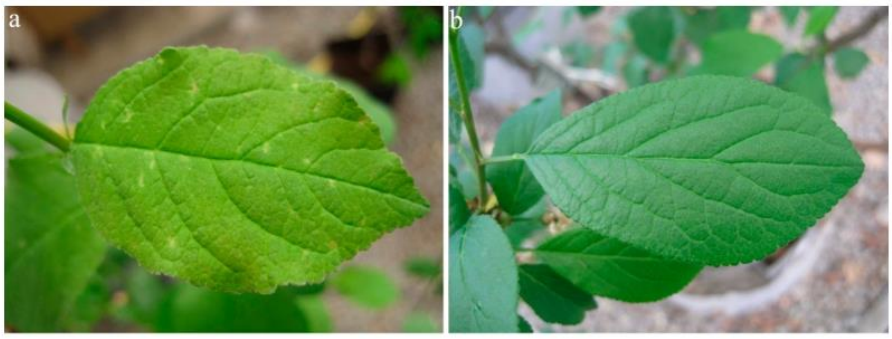

Figure 5. PNRSV symptoms on myrobalan MRS—mild spots on the leaf (**a**) compared to a healthy plant (**b**).

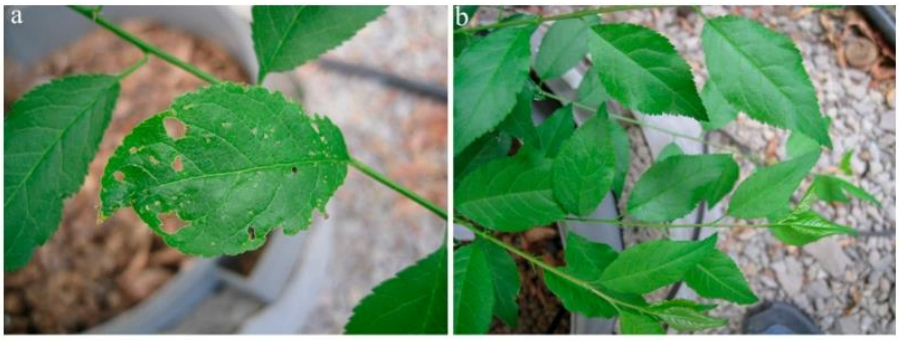

Figure 6. PNRSV symptoms on myrobalan M29C—tissue of ringspot had necrotised and dropped off (**a**) compared to a healthy plant (**b**).

Symptoms appeared in spring and in about half of the number of shoots of the plant. They occurred in a part of the shoot, not in every leaf. Shoots grown during the summer showed no symptoms.

Plants with mild symptoms or necrotic spots showed the same type of symptoms for the four years (2019–2022). However, in 2023, only mild mosaic symptoms were observed, and no necrotic ring spots were observed in both myrobalan genotypes. Apricot seedlings remained symptomless (0/10) during the five years of observation (Figure 7).

Figure 7. Apricot seedling inoculated with PNRSV, no symptoms.

4. Discussion

Two of the most common viruses infecting plants of the genus *Prunus* occurring in the Czech Republic are CVA and PNRSV [1], where the latter virus is known to cause significant economic losses [26,27], especially since plants of the genus *Prunus* (sour cherry, sweet cherry, plum, apricot, peach, etc.) in the Czech Republic play an important role in the country's economy.

Therefore, as a control measure against viruses infecting fruit trees, their propagating material is obliged to be tested for the presence of harmful viruses according to a Czech law (Act No. 219/2003 Coll. on the marketing of seeds and seedlings of cultivated plants and Decree No. 96/2018 Coll. on propagating plants and propagating material of fruit genera and species and their marketing). According to the abovementioned law, PNRSV is required to be tested in propagating material of sour cherry, sweet cherry, apricot, almond, peach, and plum.

Viruses infecting the genus *Prunus* have been monitored in Czech orchards [28], wild growing plants, including road trees of plums and myrobalans [29], and germplasm collections [1]. These studies confirmed that PNRSV and CVA are among the most common viruses infecting this genus in the Czech Republic [1].

In the current work, HTS analysis of symptomatic plants of BN4Kr myrobalan (*Prunus cerasifera* Ehrh.) revealed that these two viruses are present in a mixed infection, and the complete sequences of these viruses were obtained. The phylogenetic tree constructed on the basis of the complete coding region of RNA3 of PNRSV (Figure 1a) and using the NJ method showed that each reference isolate clustered in its appropriate molecular group, with most PNRSV isolates clustering in two groups, PV32 and PV96, in agreement with previous reports that PV32 and PV96 are the major groups. In comparison, the other four groups (PE5, CH30, SW6, and PchMx-Azt) were considered to be minor groups [30]. The Czech isolate clustered separately as a sister clade to the PV96 molecular group. Screening for SNPs leading to non-synonymous substitutions between the Czech isolate and members of other groups showed that it differed at one site in the MP and at another site in the CP. To test whether these differences might affect the virus life cycle, a search for putative targets

of PTMs was performed using ScanProsite and revealed that one site (135) in the CP, which is unique solely to the Czech isolate, was a putative target for N-myristoylation. Although there has not been enough research into how myristoylation affects the life cycle of viruses, some studies have shown that this type of PTM can be involved in membrane targeting and binding [31] and can affect the structural and functional properties of proteins, such as stabilising the conformation (spatial structure) of the protein [32]. This analysis was extended to the P1 and P2 regions and showed that the Czech isolate shared more unique amino acids with subclade II-2 (n = 28) than with clade I (n = 5) and had four unique amino acids specific to it in both regions. The search for PTMs in P1 revealed that one site was a putative target for protein kinase C phosphorylation, four sites were putative targets for casein kinase II phosphorylation, and one site was a putative target for N-myristoylation, whereas, in P2, two sites were predicted to be targets for casein kinase II phosphorylation and protein kinase C phosphorylation, respectively. Although the difference between these molecular groups cannot be related to the severity of symptoms, as the sequences used for this analysis were retrieved from NCBI GenBank and have not previously been used for comparison between P1 and P2, a previous study in 2013 by Cui and co-authors [33] found that the C-terminal of RNA1 and the 2M region of RNA2 are required for severe virulence and high levels of viral accumulation. In our report, the highest number of sites found to be targeted by PTMs was in the replicase, where two sites, 753 and 959 in P1, corresponding to positions 2258 and 2875 in the C-terminal of RNA1, were targeted for protein kinase C phosphorylation and casein kinase II phosphorylation, respectively. This observation suggests that further studies are needed to determine the potential effects of these PTMs on the viral life cycle, particularly as some previous studies have suggested that phosphorylation of the replicase may affect the function of the protein. A study by Shapka et al. [34] showed that the phosphorylation of the *cucumber necrosis tombusvirus* (CNV) P33 replication protein renders it non-functional. Phosphorylation may also be involved in replication stability; for example, the phosphorylation of the 2a protein inhibits its interaction with the 1a protein of the *cucumber mosaic virus* (CMV) [35].

Two phylogenetic trees were constructed using the complete coding regions of RNA1 and RNA2 (Figures 1a and 2a) and showed that the NCBI PNRSV isolates were divided into two major clades, I and II. Clade II was subdivided into two subclades, II-1, which included the Czech isolate, and subclade II-2. The result was similar to the previously reported phylogenetic trees constructed on the basis of the complete genome of RNA1 and RNA2 [36,37]. However, the difference between the previous studies and the present report is the observation of a new subclade II-1 containing the Czech PNRSV isolate. Kinoti and his co-authors, in 2017 [37], in their analysis of the intra-host genetic diversity of PNRSV, proposed a demarcation threshold for pairwise nucleotide similarities to distinguish different molecular groups (<97%). The pairwise nucleotide similarities (Figures 1b, 2b and 3b) between the Czech isolates and other members fell under this criterion, making the Czech PNRSV isolate a member of a new subclade, II-1, being a distant isolate from other molecular groups.

Graft testing was performed to determine the response of this divergent PNRSV isolate infection on rootstocks. Three rootstocks, Myrobalan M29C (*P. cerasifera*), MRS 2/5 (*P. cerasifera* × *P. spinosa*), and Apricot Seedling (M-VA-1), commonly used in the Czech Republic for their resistance to frost, root-knot nematodes, and their ability to improve fruit weight, were used for this purpose. The apricot seedling (M-VA-1) and myrobalan MRS 2/5 plants showed greater tolerance to the infection by being symptomless or showing mild spots, respectively. Myrobalan M29C plants showed severe symptoms with necrotic spots. Myrobalan M29C can be used as a biological indicator for the virus. Previous studies have established a chronology of the appearance of PNRSV symptoms, consisting of an acute or shock stage, one year after infection, after which, depending on the virus strain, the infected plant may either become symptomless or show recurrent symptoms annually [9,10]. In the case of this isolate, myrobalan M29C plants showed a mixture of these two patterns, starting with an acute stage, with annual recurrence of symptoms. However, after five years, the severe symptoms become mild. It is worth noting that the inoculum had CVA in

addition to PNRSV, although CVA is thought to cause latent infection [12]; however, some synergistic effects affecting the severity of symptoms can be produced in the case of mixed infections [15].

The second virus molecularly characterised in this report is the CVA co-infecting the same plant, in which the phylogenetic tree constructed using all available and non-recombinant NCBI-retrieved CVA sequences showed the presence of eight molecular groups with no particular correlation to host plants. Previously, phylogenetic analysis of CVA resulted in five phylogroups based on the RdRp region [15], and six phylogroups were proposed based on the whole genome [14]. A later study showed that according to a phylogenetic tree belt based on the complete genome of 86 sequences, the CVA population could be divided into seven phylogroups, with five divergent isolates classified as ungrouped and likely to form other groups in the presence of more isolates [16], which is the case here—the eighth observed group was formed by one of these divergent isolates together with two newly available NCBI isolates. The Czech CVA isolate (ON088603) belongs to subgroup III-2 together with other isolates from non-cherry hosts. Recombination analysis revealed 35 recombination events in 19 NCBI isolates: 8 isolates from non-cherry hosts (*Prunus serrulata* Lindl., *Prunus mume* (Siebold) Siebold & Zucc., *Prunus cerasifera* Ehrh. and *Prunus armeniaca* L.) were inferred as parental sequences, with the Czech isolate (ON088603) identified as the parental sequence of two NCBI sequences.

Recombination can be a source of viral evolution, sequence diversity, and acquisition of new hosts [38]. An interesting observation was the presence of multiple recombination events in seven isolates, which usually indicates the presence of numerous viral isolates in the same plant, since recombination requires replication of two parental genomes in the same cell [39].

This work can contribute to the general knowledge of variability in PNRSV and CVA. Although phylogroups of the two viruses cannot be linked to the geographical origin of the virus isolates, some relationship with the Asian origin of the Czech sequences and, thus, the BN4Kr myrobalan genotype can be seen. International exchange of plant breeding material is the source of the introduction of new viruses or at least new phylogroups of existing viruses into new areas. Therefore, attention to phytosanitary measures in international trade and certification of plant-propagating material is still needed to control plant viruses.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/v15081723/s1, Figure S1: Phylogenetic networks examining 105 CVA complete genomes, created by SplitsTree v.4.17.2; Figure S2: Pairwise identity matrix of the complete sequences of CVA, created by SDT v1.2 software; Table S1: List of primers; Table S2: Accession numbers of NCBI-retrieved CVA sequences; Table S3: Accession numbers of NCBI-retrieved PNRSV sequences; Table S4: Recombination events detected in CVA and PNRSV.

Author Contributions: Conceptualization, P.K. and K.B.M.; methodology, P.K., K.B.M. and J.B.; software, P.K., K.B.M. and M.K.; validation, M.K. and K.B.M.; formal analysis, K.B.M.; investigation, P.K., M.K. and J.B.; resources, P.K. and J.B.; data curation, P.K. and K.B.M.; writing—original draft preparation, P.K.; writing—review and editing, P.K. and K.B.M.; visualization, K.B.M. and M.K.; supervision, P.K.; project administration, P.K.; funding acquisition, P.K. All authors have read and agreed to the published version of the manuscript.

Funding: The work performed at CRI was supported by the Ministry of Agriculture of the Czech Republic, institutional support MZE-RO0423.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Virus genomic sequences obtained in the present work have been deposited in the GenBank of the National Center for Biotechnology Information (NCBI) under accession numbers ON088600, ON088601, ON088602 and ON088603.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Přibylová, J.; Lenz, O.; Fránová, J.; Koloniuk, I.; Špak, J. Comparison of the Incidence of Viruses within Germplasm, Orchards and Wild Cherry Trees in the Czech Republic. *Ann. Appl. Biol.* **2020**, *176*, 138–146. [CrossRef]
2. Pallas, V.; Aparicio, F.; Herranz, M.C.; Sanchez-Navarro, J.A.; Scott, S.W. The Molecular Biology of Ilarviruses. In *Advances in Virus Research*; Elsevier Inc.: Amsterdam, The Netherlands, 2013; Volume 87, pp. 139–181. ISBN 9780124076983.
3. Glasa, M.; Betinová, E.; Kúdela, O.; Šubr, Z. Biological and Molecular Characterization of Prunus Necrotic Ringspot Virus Iolates and Possible Approaches to Their Phylogenetic Typing. *Ann. Appl. Biol.* **2002**, *140*, 279–283. [CrossRef]
4. Boulila, M. Putative Recombination Events and Evolutionary History of Five Economically Important Viruses of Fruit Trees Based on Coat Protein-Encoding Gene Sequence Analysis. *Biochem. Genet.* **2010**, *48*, 357–375. [CrossRef] [PubMed]
5. Song, S.; Sun, P.-P.; Chen, Y.-H.; Ma, Q.; Wang, X.-Y.; Zhao, M.-M.; Li, Z.-N. Complete Genome Sequences of Five Prunus Necrotic Ringspot Virus Isolates from Inner Mongolia of China and Comparison to Other PNRSV Isolates around the World. *J. Plant Pathol.* **2019**, *101*, 1047–1054. [CrossRef]
6. Karešová, R.; Navrátil, M.; Pluhař, Z.; Paprštein, F. First Results from Using the ELISA Method for Detection of Prunus Necrotic Ringspot Virus in Sweet Cherry, Sour Cherry and Plums. *Hortic. Sci.* **1986**, *3*, 169–176.
7. Kryczynski, S.; Szyndel, M.S.; Stawiszynska, A.; Piskorek, W. The Rate and the Way of Prunus Necrotic Ringspot Virus Spread in Sour Cherry Orchard and in the Rootstock Production. *Acta Hortic.* **1992**, *309*, 105–110. [CrossRef]
8. Barba, M.; Pasquini, G.; Quacquarelli, A. Roles of Seeds in the Epidemiology of Two Almond Viruses. *Acta Hortic.* **1986**, *193*, 127–130. [CrossRef]
9. Nyland, G.; Gilmer, R.M.; Moore, J.D. "Prunus" Ringspot Group. In *Virus Diseases and Noninfectious Disorders of Stone Fruits in North America*; Agriculture Handbook No. 437; USDA: Washington, DC, USA, 1976.
10. Wells, J.M.; Kirkpatrick, H.C. Symptomatology and Incidence of Prunus Necrotic Ringspot Virus in Peach Orchards in Georgia. *Plant Dis.* **1986**, *70*, 444–447. [CrossRef]
11. Jelkmann, W. Cherry Virus A: CDNA Cloning of DsRNA, Nucleotide Sequence Analysis and Serology Reveal a New Plant Capillovirus in Sweet Cherry. *J. Gen. Virol.* **1995**, *76*, 2015–2024. [CrossRef] [PubMed]
12. Gao, R.; Li, S.; Lu, M. Complete Nucleotide Sequences of Two Isolates of Cherry Virus A from Sweet Cherry in China. *J. Integr. Agric.* **2016**, *15*, 1667–1671. [CrossRef]
13. Grimová, L.; Zouhar, M.; Ryšánek, P.; Drabešová, J.; Mazáková, J.; Paprštein, F. First Occurence of Cherry Virus A (CVA) in the Czech Republic. *Jul.-Kühn-Arch.* **2010**, *427*, 2009–2011.
14. Kesanakurti, P.; Belton, M.; Saeed, H.; Rast, H.; Boyes, I.; Rott, M. Comparative Analysis of Cherry Virus A Genome Sequences Assembled from Deep Sequencing Data. *Arch. Virol.* **2017**, *162*, 2821–2828. [CrossRef]
15. Marais, A.; Svanella-Dumas, L.; Barone, M.; Gentit, P.; Faure, C.; Charlot, G.; Ragozzino, A.; Candresse, T. Development of a Polyvalent RT-PCR Detection Assay Covering the Genetic Diversity of *Cherry Capillovirus* A. *Plant Pathol.* **2012**, *61*, 195–204. [CrossRef]
16. Gao, R.; Xu, Y.; Candresse, T.; He, Z.; Li, S.; Ma, Y.; Lu, M. Further Insight into Genetic Variation and Haplotype Diversity of Cherry Virus A from China. *PLoS ONE* **2017**, *12*, e0186273. [CrossRef] [PubMed]
17. Polák, J.; Komínek, P. Evaluation of Rootstocks of Stone Fruits for Resistance to Natural Plum Pox Virus Infection. *Can. J. Plant Pathol.* **2014**, *36*, 116–120. [CrossRef]
18. Komínek, P.; Massart, S.; Pham, K.; van Leeuwen, P.; Komínková, M. Characterisation of a Novel Virus Infecting Orchids of the Genus Pleione. *Virus Res.* **2019**, *261*, 56–59. [CrossRef] [PubMed]
19. Alzohairy, A.M. BioEdit: An Important Software for Molecular Biology. *GERF Bull. Biosci.* **2011**, *2*, 60–61.
20. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT Online Service: Multiple Sequence Alignment, Interactive Sequence Choice and Visualization. *Brief. Bioinform.* **2018**, *20*, 1160–1166. [CrossRef] [PubMed]
21. Abascal, F.; Zardoya, R.; Telford, M.J. TranslatorX: Multiple Alignment of Nucleotide Sequences Guided by Amino Acid Translations. *Nucleic Acids Res.* **2010**, *38*, W7–W13. [CrossRef]
22. Martin, D.P.; Murrell, B.; Golden, M.; Khoosal, A.; Muhire, B. RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes. *Virus Evol.* **2015**, *1*, vev003. [CrossRef] [PubMed]
23. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef] [PubMed]
24. Muhire, B.M.; Varsani, A.; Martin, D.P. SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS ONE* **2014**, *9*, e108277. [CrossRef]
25. Huson, D.H.; Bryant, D. Application of Phylogenetic Networks in Evolutionary Studies. *Mol. Biol. Evol.* **2006**, *23*, 254–267. [CrossRef] [PubMed]
26. Zong, X.; Wang, W.; Wei, H.; Wang, J.; Yan, X.; Hammond, R.W.; Liu, Q. Incidence of Sweet Cherry Viruses in Shandong Province, China and a Case Study on Multiple Infection with Five Viruses. *J. Plant Pathol.* **2015**, *97*, 61–68. [CrossRef]
27. Sokhandan-Bashir, N.; Kashiha, Z.; Koolivand, D.; Eini, O. Detection and Phylogenetic Analysis of Prunus Necrotic Ringspot Virus Isolates from Stone Fruits in Iran. *J. Plant Pathol.* **2017**, *99*, 723–729.
28. Suchá, J.; Svobodová, L. Incidence of Prune Dwarf Virus and Prunus Necrotic Ring Spot Virus in Orchards of Sweet and Sour Cherry in the Czech Republic. *Hortic. Sci.* **2010**, *37*, 118–120. [CrossRef]

29. Polák, J. Viruses of Blackthorn and Road-Bordering Trees of Plum, Myrobalan, Sweet and Sour Cherries in the Czech Republic. *Plant Prot. Sci.* **2007**, *43*, 1–4. [CrossRef]
30. Pallas, V.; Aparicio, F.; Herranz, M.C.; Amari, K.; Sanchez-Pina, M.A.; Myrta, A.; Sanchez-Navarro, J.A. Ilarviruses of Prunus Spp.: A Continued Concern for Fruit Trees. *Phytopathology* **2012**, *102*, 1108–1120. [CrossRef] [PubMed]
31. Strecker, T.; Maisa, A.; Daffis, S.; Eichler, R.; Lenz, O.; Garten, W. The Role of Myristoylation in the Membrane Association of the Lassa Virus Matrix Protein Z. *Virol. J.* **2006**, *3*, 93. [CrossRef]
32. Resh, M.D. Fatty Acylation of Proteins: New Insights into Membrane Targeting of Myristoylated and Palmitoylated Proteins. *Biochim. Biophys. Acta* **1999**, *1451*, 1–16. [CrossRef]
33. Cui, H.; Hong, N.; Wang, G.; Wang, A. Genomic Segments RNA1 and RNA2 of Prunus Necrotic Ringspot Virus Codetermine Viral Pathogenicity to Adapt to Alternating Natural Prunus Hosts. *Mol. Plant-Microbe Interact.* **2013**, *26*, 515–527. [CrossRef] [PubMed]
34. Shapka, N.; Stork, J.; Nagy, P.D. Phosphorylation of the P33 Replication Protein of Cucumber Necrosis Tombusvirus Adjacent to the RNA Binding Site Affects Viral RNA Replication. *Virology* **2005**, *343*, 65–78. [CrossRef] [PubMed]
35. Kim, S.H.; Palukaitis, P.; Park, Y.I. Phosphorylation of Cucumber Mosaic Virus RNA Polymerase 2a Protein Inhibits Formation of Replicase Complex. *EMBO J.* **2002**, *21*, 2292–2300. [CrossRef] [PubMed]
36. Noorani, M.S.; Baig, M.S.; Khan, J.A.; Pravej, A. Whole Genome Characterization and Diagnostics of Prunus Necrotic Ringspot Virus (PNRSV) Infecting Apricot in India. *Sci. Rep.* **2023**, *13*, 4393. [CrossRef] [PubMed]
37. Kinoti, W.M.; Constable, F.E.; Nancarrow, N.; Plummer, K.M.; Rodoni, B. Analysis of Intra-Host Genetic Diversity of Prunus Necrotic Ringspot Virus (PNRSV) Using Amplicon next Generation Sequencing. *PLoS ONE* **2017**, *12*, e0179284. [CrossRef]
38. García-Arenal, F.; Fraile, A.; Malpica, J.M. Variation and Evolution of Plant Virus Populations. *Int. Microbiol.* **2003**, *6*, 225–232. [CrossRef] [PubMed]
39. Morilla, G.; Krenz, B.; Jeske, H.; Bejarano, E.R.; Wege, C. Tête à Tête of Tomato Yellow Leaf Curl Virus and Tomato Yellow Leaf Curl Sardinia Virus in Single Nuclei. *J. Virol.* **2004**, *78*, 10715–10723. [CrossRef]

Chapter 4: Summary discussion

The objective of this doctoral thesis was to utilise a range of bioinformatic tools to elucidate the evolutionary dynamics and genetic characterization of several plant viruses. This doctoral thesis comprises four distinct studies, the results of which have already been published. Two studies were done with the objective of elucidating the origin and evolutionary history of WMV (study 1) and GPGV (study 2), which have population with different proportions of recombinant isolates. These frequencies influence the temporal signal (i.e. affect the ability to estimate dates) and therefore different methods were used to date their phylogenies.

In study 1, the Czech WMV population was found to be a closely related cluster within the world population. This population exhibited the lowest genetic diversity ($\pi = 0.010 \pm 0.00013$) when compared to the other populations tested. This indicates a lower level of evolutionary divergence when compared to populations from China, South Korea, France, Italy, and the United States. The recombination analysis of WMV revealed that the frequency of recombinant WMV isolates was 96.24%, with the majority of putative parents originating from Asia. There were two types of intraspecific recombination that were found: simple (36%) and multiple (61%), and the most affected region involved parts of the P1 coding region. The study demonstrated that WMV is a highly recombinogenic potyvirus, exhibiting greater recombination potential than any other studied potyvirus (Ben Mansour *et al.*, 2023). In light of this information, the strategy of subtree dating method as employed by Mohammadi *et al.*, (2018) and Fuentes *et al.*, (2019, 2021) was adopted in study 1 to estimate the dates of nodes in WMV phylogeny. The study revealed that the earliest population of WMV infected a number of plant species including *Ailanthus altissima*, *Alcea rosea*, and *Panax ginseng*. These isolates were phylogenetically distinct from other WMV isolates. Subtree dating indicated that the virus first appeared in north China at least 2000 years ago in non-cucurbit hosts, and subsequently, migrated to watermelon, which was first grown as a crop in northern China more than 1000 years ago. This was therefore most likely a 'new encounter spillover' (Gibbs *et al.*, 2008), and subsequently WMV became adapted to watermelon spreading in seed of favoured watermelon varieties as they were spread worldwide.

In Study 2, the GPGV isolates were divided into five statistically distinct clusters designated A-E. The German GPGV isolates were dispersed into four of the clusters, indicating that they are genetically more diverse than the GPGV populations of other European populations. The results of

the analysis of evolutionary selection pressure ratio indicate that both the MP and CP coding regions of the German GPGV isolates were under negative selection pressure. This is a purifying process that removes isolates with deleterious mutations and decreases the frequency of less-fit viral variants (García-Arenal *et al.*, 2003; Hughes, 2009). The investigation of the site-specific selection pressure affecting the MP and CP genes of the German population revealed the presence of only six codon sites that were assigned for a positive selection pressure. These sites were located at the located at the 3′ end of the MP and at one codon site for the CP. It is noteworthy that sites under positive selection were present more often in the MP than in the CP. The positive selection of the MP gene may facilitate the spread of beneficial virus variants by promoting systemic infection (Yoshikawa *et al.*, 2006; Zhang, 2008). The recombination analysis revealed that only three GPGV isolates were recombinant. These isolates were removed, and the remaining sequences demonstrated a significant temporal signal ($p = 0.01–0.001$) and estimated a most recent common ancestor (MRCA) dated around 1285 CE. However, since some of the NCBI-retrieved sequences lacked any mention of their collection date, a time-scaled tree was calculated using IQ-Tree v.2.2.2. The molecular dating enabled updated of the likely origin of GPGV, indicating that it diverged from grapevine inner necrosis virus (GINV) from wild grape (*Vitis coignetiae*) probably around 3500 years ago in North East Asia or Japan. Subsequently, GPGV infected the Eurasian grape when those cultivars were first taken to China around 2400 years ago. The introduction of GPGV to Europe was close to 1800 CE. The German isolates were only identified in all parts of the post-1800 CE phylogeny. The genetic diversity exhibited by the German isolates was greater than that observed in other European populations, as evidenced by the results obtained from both the MP and CP genes. This suggests that the initial stages of the GPGV invasion of Europe occurred in Germany, rather than Italy (Ben Mansour *et al.*, 2024) as previously suggested (Hily *et al.*, 2020).

Additionally, two studies were conducted to investigate the prevalence and genetic characterisation of viruses found in two plants, namely the grapevine (study 3), and the myrobalan plum (study 4).

In Study 3, samples from asymptomatic grapevines were subjected to HTS analysis. The full-length sequences of four grapevine viruses GFkV, GRGV, GRVFV, and GSyV-1 were successfully obtained, along with other partial genomic parts. GAMaV was identified for the first time in the Czech Republic. The phylogenetic analysis revealed the presence of both mixed infections of different viruses of the family *Tymoviridae* and distinct molecular variants of the same virus within

the same grapevine. A high divergence in ORF3 and ORF4 at the amino acid level was observed between the Czech and Italian GFkV isolates. Evidence of intraspecific recombination was identified in the Czech isolates of GRVFV, GSyV-1 and GRGV isolates was found. It was intriguing that all identified recombinants of GRVFV had either one or both parental sequences from a non-*Vitis* host (*Prunus*), as indicated by the information in the NCBI database. Although, this information should be treated with caution, it suggests that recombination may occur between GRVFV isolates from *Vitis* and non-*Vitis* hosts. It has been proposed that an insect vector is responsible for field transmission due to the presence of multiple infections of GRVFV+GRGV within the same plant (Cretazzo & Velasco, 2016). The presence of recombinants with parental sequences from non-*Vitis* hosts suggests the probable involvement of polyphagous insects in the transmission of GRVFV. A second noteworthy observation is the evidence of multiple recombination events for GSyV-1 isolates from central Europe. This may indicate that the population is genetically diverse because the isolates recombine freely (Zakubanskiy *et al.*, 2018), and this is consistent with a Hungarian study, which demonstrated that isolates from central European isolates are genetically diverse (Czotter *et al.*, 2018).

In Study 4, HTS analysis of symptomatic plants of BN4Kr myrobalan (*Prunus cerasifera* Ehrh.) revealed the presence of two viruses (PNRSV and CVA) in a mixed infection. The Czech PNRSV isolate was found to cluster separately as a sister clade to the PV96 molecular group. Screening for single nucleotide polymorphisms (SNPs) leading to non-synonymous substitutions between the Czech isolate and members of other groups showed that it differed at one site in the MP and at another site in the CP. In 2017, Kinoti and his colleagues, proposed a demarcation threshold for pairwise nucleotide similarities to distinguish different molecular groups of PNRSV (Kinoti *et al*, 2017). This threshold was set at a value of less than 97% and using this criterion placed the Czech PNRSV isolate as distinct from other groups of PNRSV. Graft testing of this divergent PNRSV isolate infection on rootstocks was done. Three rootstocks were employed for this purpose: myrobalan M29C (*P. cerasifera*), MRS 2/5 (*P. cerasifera* × *P. spinosa*), and apricot seedling (M-VA-1). The apricot seedling (M-VA-1) and myrobalan MRS 2/5 plants showed greater tolerance to the infection, evidence by the absence of symptoms or the presence of mild spots, respectively. Myrobalan M29C plants exhibited severe symptoms, including necrotic spots. The myrobalan M29C plant can be used as a biological indicator for the virus.

The Czech CVA isolate was found to be phylogenetically related to other isolates from non-cherry hosts. A recombination analysis revealed 35 recombination events. Eight isolates from non-cherry hosts (*Prunus serrulata* Lindl., *Prunus mume* (Siebold) Siebold & Zucc., *Prunus cerasifera* Ehrh. and *Prunus armeniaca* L.) were identified as parental sequences, with the Czech isolate identified as the parental sequence of two NCBI sequences. Recombination can be a source of viral evolution, sequence diversity, and the acquisition of new hosts (García-Arenal *et al.*, 2003). An interesting observation was the presence of multiple recombination events in seven isolates, which typically indicates the presence of numerous viral isolates in the same plant.

Conclusions

The hypothesis that the adaptation of plant viruses to new host species can be elucidated by examining historical evidence was validated. The result of our first study indicates that WMV originated in East Asia more than 2000 years ago. Prior to its global dispersal, the virus first infected non-crop hosts and subsequently transferred to a newly imported crop approximately 1000 years ago. This is consistent with previous studies that have suggested that BCMV lineage, of which WMV is a member, emerged more than 3000 years ago in Asia.

Similarly, a previous study indicated that GPGV originated from Asia, with China as a probable origin for the virus emerged. Although, our study, confirmed that the virus originated in Asia, but probably further north and east, possibly Japan or Korea, and that the virus diverged from GINV in wild grapevine before moving to Eurasian grapevine.

The second hypothesis suggesting that the viruses infecting trees exhibit distinct genetic characteristics and distribution patterns was also validated. The examined grapevine plants showed the presence of mixed infections, as well as several variants of these viruses in the same plant. And that based on recombination pattern, a potential involvement of polyphagous insects in the transmission of GRVFV.

The identified PNRSV isolate was phylogenetically distinct from other isolates, and a biological assay demonstrated that this isolate induces different symptom patterns than the common one.

In conclusion, the current findings offer valuable insights into the genetic diversity, distribution, origins, and probable transmission manners of these viruses.

List of cited references (outside of published papers)

Ágoston J, Almási A, Salánki K, Palkovics L, 2020. Genetic diversity of potyviruses associated with tulip breaking syndrome. *Plants* **9**, 1–25.

Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P, 2004. Emerging infectious diseases of plants: Pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology and Evolution* **19**, 535–544.

Anon, 2024a. Charles Darwin. Available from: https://en.wikipedia.org/wiki/Charles_Darwin [Accessed 13 May 2024].

Anon, 2024b. Alfred Russel Wallace. Available from: https:// en.wikipedia.org/wiki/ Alfred_Russel_Wallace [Accessed 13 May 2024].

Bawono P, Heringa J, 2014. Phylogenetic Analyses. In Brahme A, Ed. *Comprehensive Biomedical Physics*. Elsevier UK. 93-110

Ben Mansour K, Gibbs AJ, Komínková M, Komínek P, Brožová J, Kazda J, Zouhar M, Ryšánek P, 2023. Watermelon mosaic virus in the Czech Republic, its recent and historical origins. *Plant Pathology* **72**, 1528–1538.

Ben Mansour K, Gibbs AJ, Meßmer N, Fuchs R, Wetzel T, Winterhagen P, 2024. Grapevine Pinot gris virus in Germany: From where did the virus come, and when? *Plant Pathology* **73**, 455–464.

Boscia D, Martelli GP, Savino V, Castellano MA, 1991. Identification of the agent of grapevine fleck disease. *Vitis* **30**, 97–105.

Burrell CJ, Howard CR, Murphy FA, 2017. Classification of viruses and phylogenetic relationships. *Fenner and White's Medical Virology*, 15–25.

Castellano M, Martelli G, Savino V, 1983. Virus-like particles and ultrastructural modifications in the phloem of leafroll-afected grapevines. *Vitis* **22**, 23–39.

Chambers PJ, Pretorius IS, 2010. Fermenting knowledge: The history of winemaking, science and yeast research. EMBO Rep **11**, 914-920.

Chomicki G, Schaefer H, Renner SS, 2020. Origin and domestication of cucurbitaceae crops: insights from phylogenies, genomics and archaeology. *New Phytologist* **226**, 1240–1255.

Choudhuri S, 2014. Phylogenetic Analysis. In: *Bioinformatics for Beginners: Genes, genomes, molecular evolution, databases and analytical tools*. Elsevier Academic Press. 209–218,

Czortek P, Adamowski W, Kamionka-Kanclerska K, Karpińska O, Zalewski A, Dyderski MK, 2024. Patterns of Prunus cerasifera early invasion stages into a temperate primeval forest. *Biological Invasions* **26**, 633–647.

Drake JW, Holland JJ, 1999. Mutation rates among RNA viruses. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 13910–13913.

Escriu F, 2017. Diversity of plant virus populations: A valuable tool for epidemiological studies. In: Bitz L, ed. *Genetic diversity*. Intechopen, London, 160.

Ferriol M, Picó B, 2008. Pumpkin and Winter Squash. In: Prohens J,, Nuez F, eds. *Handbook of Plant Breeding Vegetables I Part 4, Volume 1*. Springer, Berlin, 317–349.

Fraile A, Escriu F, Aranda MA, Malpica JM, Gibbs AJ, García-Arenal F, 1997. A century of tobamovirus evolution in an Australian population of Nicotiana glauca. *Journal of Virology* **71**, 8316–8320.

French R, Stenger DC, 2003. Evolution of wheat streak mosaic virus: Dynamics of population growth within plants may explain limited variation. *Annual Review of Phytopathology* **41**, 199–214.

Gergerich RC, Dolja VV., 2006. Introduction to plant viruses, the invisible foe. *The Plant Health Instructor*.

Gibbs A, 2020. Binomial nomenclature for virus species: a long view. *Archives of Virology* **165**, 3079–3083.

González-Neves G, Gil G, Favre G, Baldi C, Hernández N, Traverso S, 2013. Influence of winemaking procedure and grape variety on the colour and composition of young red wines. *South African Journal of Enology and Viticulture* **34**, 138–146.

Gorbalenya AE, Lauber C, 2017. Phylogeny of Viruses. In: *Reference Module in Biomedical Sciences*. Elsevier.

Gospodaryk A, Moročko-Bičevska I, Pupola N, Kale A, 2013. Occurrence of stone fruit viruses in plum orchards in Latvia. *Proceedings of the Latvian Academy of Sciences, Section B: Natural, Exact, and Applied Sciences* **67**, 116–123.

Green KJ, Brown CJ, Gray SM, Karasev A V., 2017. Phylogenetic study of recombinant strains of Potato virus Y. *Virology* **507**, 40–52.

Guo J, Xu W, Hu Y *et al.*, 2020. Phylotranscriptomics in cucurbitaceae reveal multiple whole-genome duplications and key morphological and molecular innovations. *Molecular Plant* **13**, 1117–1133.

Hughes A., 2009. Small effective population sizes and rare nonsynonymous variants in potyviruses. *Virology* **393**, 127–134.

Jo Y, Choi H, Kyong Cho J, Yoon JY, Choi SK, Kyong Cho W, 2015. In silico approach to reveal viral populations in grapevine cultivar Tannat using transcriptome data. *Scientific Reports* **5**, 1–11.

Jones RAC, Naidu RA, 2019. Global Dimensions of Plant Virus Diseases: Current status and future perspectives. *Annual Review of Virology* **6**, 387–409.

Jones RAC, Vazquez-Iglesias I, Hajizadeh M, McGreig S, Fox A, Gibbs AJ, 2022. Phylogenetics and evolution of wheat streak mosaic virus: Its global origin and the source of the Australian epidemic. *Plant Pathology* **71**, 1660–1673.

Kamenova I., 2008. Prunus cerasifera as a host of plum pox virus in Bulgaria. *Journal of Plant Pathology* **90**, S15–S18.

Khanal V, Ali A, 2021. High mutation frequency and significant population differentiation in

papaya ringspot virus-w isolates. *Pathogens* **10**.

LaTourrette K, Garcia-Ruiz H, 2022. Determinants of virus variation, evolution, and host adaptation. *Pathogens* **11**, 1039.

Lecoq H, Desbiez C, 2012. Viruses of cucurbit crops in the Mediterranean region: An ever-changing picture. In: *Advances in Virus Research*. 67–126.

MacFarlane SA, 1997. Natural recombination among plant virus genomes: Evidence from Tobraviruses. *Seminars in Virology* **8**, 25–31.

Maliogka VI, Salvador B, Carbonell A *et al.*, 2012. Virus variants with differences in the p1 protein coexist in a plum pox virus population and display particular host-dependent pathogenicity features. *Molecular Plant Pathology* **13**, 877–886.

Malpica JM, Fraile A, Moreno I, Obies CI, Drake JW, García-Arenal F, 2002. The rate and character of spontaneous mutation in RNA virus. *Genetics* **162**, 1505–1511.

Maynard Smith J, 1998. *Evolutionary Genetics*. Oxford, UK.

Milliron L, Niederholzer F, Buchner R *et al.*, 2021. Field evaluation of rootstocks for "improved French" prune production in California grower trials. *Acta Horticulturae* **1322**, 125–130.

Mitrofanova I V., Zakubanskiy A V., Mitrofanova O V., 2018. Viruses infecting main ornamental plants: An overview. *Ornamental Horticulture* **24**, 95–102.

Mohammadi M, Gibbs AJ, Hosseini A, Hosseini S, 2018. An Iranian genomic sequence of Beet mosaic virus provides insights into diversity and evolution of the world population. *Virus Genes* **54**, 272–279.

Morata A, Loira I, Vejarano R, González C, Callejo MJ, Suárez-Lepe JA, 2017. Emerging preservation technologies in grapes for winemaking. *Trends in Food Science and Technology* **67**, 36–43.

Moury B, Desbiez C, Jacquemond M, Lecoq H, 2006. Genetic diversity of plant virus populations: Towards hypothesis testing in molecular epidemiology. *Advances in Virus Research* **67**, 49–87.

Nagy PD, 2008. Recombination in plant RNA viruses. *Plant Virus Evolution*, 133–156.

Ouedraogo RS, Roossinck MJ, 2018. Molecular Evolution. In: Palukaitis PF, García-Arenal F, eds. *Cucumber Mosaic Virus*. APS Press, Minneapolis, MN, 207–215.

Pagán I, 2018. The diversity, evolution and epidemiology of plant viruses: A phylogenetic view. *Infection, Genetics and Evolution* **65**, 187–199.

Pallas V, Aparicio F, Herranz MC, Sanchez-Navarro JA, Scott SW, 2013. The molecular biology of Ilarviruses. In: *Advances in Virus Research*. Elsevier Inc., 139–181.

Pappas N, Roux S, Hölzer M *et al.*, 2020. Virus Bioinformatics.In: *Encyclopedia of Virology: Volume 1-5, Fourth Edition* **1**–**5**, 124–132.

Paris HS, 2016. Overview of the origins and history of the five major cucurbit crops: issues for ancient DNA analysis of archaeological specimens. *Vegetation History and Archaeobotany*

25, 405–414.

Pérez-losada M, Arenas M, Galán JC, Palero F, González-Candelas F, 2015. Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. *Infection, Genetics and Evolution* **30**, 296–307.

Petrov D, Ocokoljić M, Galečić N, Skočajić D, Simović I, 2024. Adaptability of Prunus cerasifera Ehrh. to Climate Changes in Multifunctional Landscape. *Atmosphere* **15**.

Popescu I, Caudullo G, 2016. Prunus cerasifera in Europe: Distribution, habitat, usage and threats. In: San-Miguel-Ayanz J, de Rigo D, Caudullo G, Houston Durrant T, Mauri A, eds. *European Atlas of Forest Tree Species*. Luxembourg, Publications Office of the European Union.

Provesi JG, Amante ER, 2015. *Carotenoids in Pumpkin and Impact of Processing Treatments and Storage*. Elsevier Inc.

Renner SS, Wu S, Pérez-Escobar OA, Silber M V., Fei Z, Chomicki G, 2021. A chromosome-level genome of a Kordofan melon illuminates the origin of domesticated watermelons. *Proceedings of the National Academy of Sciences of the United States of America* **118**, 1–9.

Schaefer H, Renner SS, 2011. Phylogenetic relationships in the order Cucurbitales and a new classification of the gourd family (Cucurbitaceae). *Taxon* **60**, 122–138.

Simmonds P, Adams MJ, Benk M *et al.*, 2017. Consensus statement: Virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology* **15**, 161–168.

Simmons MP, Ochoterena H, Freudenstein J V., 2002. Amino acid vs. nucleotide characters: Challenging preconceived notions. *Molecular Phylogenetics and Evolution* **24**, 78–90.

Southwick SM, Yeager JT, Osgood J, Olson W, Norton M, Buchner R, 1999. Performance of new Marianna rootstocks in California for “French” prune. *HortTechnology* **9**, 498–505.

Stevens WA, 1983. Symptoms of Plant Virus Infection. In: *Virology of Fowering Plants. Tertiary Level Biology*. Springer, Boston, MA, 5–24.

Tatineni S, Hein GL, 2023. Plant Viruses of Agricultural Importance: Current and Future Perspectives of Virus Disease Management Strategies. *Phytopathology* **113**, 117–141.

Verter N, Hasíková L, 2019. The performance of grape production in the Czech Republic. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **67**, 333–342.

Vončina D, Al Rwahnih M, Rowhani A, Gouran M, Almeida RPP, 2017. Viral diversity in autochthonous Croatian grapevine cultivars. *Plant Disease* **101**, 1230–1235.

Zhang J, 2008. Positive selection, not negative selection, in the pseudogenization of rcsA in Yersinia pestis. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 2008.

Zhu J, Parker A, Gou F *et al.*, 2021. Developing perennial fruit crop models in APSIM Next Generation using grapevine as an example. *In Silico Plants* **3**, 1–23.

Zohary D, 1992. Is the European plum, Prunus domestica L., a P. cerasifera EHRH. x P. spinosa L. allo-polyploid? *Euphytica* **60**, 75–77.

List of publications

Charon J, Olendraite I, Forgia M, Chong Li C, Hillary LS, Roux S, Kupczok A, Debat H, Sakaguchi S, Tahzima R, Nakagawa S, Babaian Artem, Abroi A, Bejerman N, **Ben Mansour K**, et al., 2024. Consensus statement from the first RdRp Summit: advancing RNA virus discovery at scale across communities. *Frontiers in Virology*, 4 :1371958. https://doi.org/10.3389/fviro.2024.1371958

Komínková M, **Ben Mansour K**, Komínek P, Brožová J, Střalková R, 2024. Multiple Infections with Viruses of the Family Tymoviridae in Czech Grapevines. *Viruses*, **16**, 343.

Ben Mansour K, Gibbs AJ, Komínková M, Komínek P, Brožová J, Kazda J, Zouhar M, Ryšánek P, 2023. Watermelon mosaic virus in the Czech Republic, its recent and historical origins. *Plant Pathology*, **72**, 1528–1538. Available from: https://doi.org/10.1111/ppa.13766

Ben Mansour K, Komínek P, Komínková M, Brožová J, 2023. Characterization of Prunus Necrotic Ringspot Virus and Cherry Virus A Infecting Myrobalan Rootstock. *Viruses*, **15** (8):1723. https://doi.org/10.3390/v15081723

Ben Mansour K, Gibbs AJ, Meßmer N, Fuchs R, Wetzel T, Winterhagen P, 2023. Grapevine Pinot gris virus in Germany: From where did the virus come, and when? *Plant Pathology*, **73**, 455–464. Available from: https://doi.org/10.1111/ppa.13824

Mathioudakis M, Khechmar S, Owen C, Medina V**, Ben Mansour K,** Tomaszewska W, et al., 2018. A Thioredoxin Domain-Containing Protein Interacts with Pepino mosaic virus Triple Gene Block Protein. Int J Mol Sci. 2018; 19:3747.

List of publications under preparation

Ben Mansour K, Kolonuik I, Přibylová J, Špak J, Sedlák J, Sarkisova T, Semerák M, Komínek P, Brožová J, Komínková M. The occurrence and variability of novel apple viruses and viroid in the Czech Republic.

Ben Mansour K, Gibbs AJ, Špak J, Kominek P, Zouhar M, Rysanek P. High-throughput sequencing of Erysimum latent virus isolate-*Sisymbrium altissimum*.

Ben Mansour K, Kominkova M, Kominek P, Kazda J, Račka J, Brožová J, Wenzlová J, Zouhar M, Rysanek P. Current status of viruses infecting pumpkin in the Czech Republic.

Gibbs AJ, **Ben Mansour K**, Hajizadeh M. Spillovers: a study of the genetic differences between crop and non-crop potyviruses of the Bean Common Mosaic subgroup.

Komínek P, Komínková M, **Ben Mansour K,** Ponert J. Analysis of virome of botanical garden with special emphasis on the family Orchidaceae.

List of active participation in conferences

Ben Mansour K, Gibbs AJ, Messmer N, Bohnert P, Fuchs R, Wetzel T, Winterhagen P. Update on the origin of grapevine Pinot gris virus and its molecular characterization in Germany. *The 20th conference of the International Council for the Study of Virus and Virus-Like Diseases of the Grapevine*-Greece. **25-29.09.2023**: p.208-209.

Ben Mansour K, Gibbs AJ, Komínková M, Komínek P, Brožová J, Kazda J, Zouhar M, Ryšánek P. Sequencing, and recombination analysis of watermelon mosaic isolates from the Czech Republic. *The 12th International Congress of Plant Pathology*-France. **20-25.08.2023**: p835-836

Ben Mansour K, Gibbs AJ, Komínková M, Komínek P, Brožová J, Kazda J, Zouhar M, Ryšánek P. Genetic characterization of watermelon mosaic virus. *International Virus Bioinformatics Meeting*-Spain. **24-26.05.2023**

Ben Mansour K, Komínková M, Komínek P, Brožová J, Kazda J, Zouhar M, Ryšánek P. Genetic Characterization of watermelon mosaic virus isolates in the Czech Republic. *The 62 sd Phylogenetic Symposium*-Germany. **18-20.11.2022**

Ben Mansour K, Messmer N, Fuchs R, Wetzel T, Rysanek P, Winterhagen P. Genetic characterization of grapevine pinot gris in Germany. *The International Advances in Plant Virology*-Slovenia. **05-07.10.2022**

Ben Mansour K, Kazda J, Grimová L, Komínková M, Komínek P, Zouhar M, Ryšánek, P. Current status of viruses infecting pumpkins fields and their occurrence in weeds in the Czech Republic. *The Czech and Slovak conference on plant protection*-Czech Republic. **08-9.09.2022**: p 15

Ben Mansour K, Messmer N, Fuchs R, Wetzel T, Ryšánek P, Winterhagen P. Molecular characterization of German Grapevine pinot gris virus isolates indicates genetic variability. *The Czech and Slovak conference on plant protection*-Czech Republic. **08-9.09.2022**: p 16

Ben Mansour K, Kazda J, Grimová L, Komínková M, Komínek P, Zouhar M, Ryšánek P. Current status of viruses infecting pumpkins fields and their occurrence in weeds in the Czech Republic. *Plant pathology, entomology, and the road ahead*. Virtual meeting. **28-30.06.2022**

Ben Mansour K, Komínková M, Komínek P, Brožová J, Kazda J, Zouhar M, Ryšánek P. Molecular characterization of watermelon mosaic virus isolates infecting zucchini and pumpkin plants in the Czech Republic. *The 15th International Plant Virus Epidemiology Symposium*-Spain. **05-08.06.2022**: p 113